



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

CNN-BiLSTM 하이브리드 모델을
활용한 심전도 부정맥 분류:
MFCC 및 Softmax 특성의
과업 의존적 유효성 평가

연세대학교 대학원

융합의학과

이 병 훈

CNN-BiLSTM 하이브리드 모델을
활용한 심전도 부정맥 분류:
MFCC 및 Softmax 특성의
과업 의존적 유효성 평가

지도교수 구성욱 · 장원석

이 논문을 석사 학위논문으로 제출함

2025년 07월

연세대학교 대학원

융합의학과

이 병 훈

CNN-BiLSTM 하이브리드 모델을 활용한 심전도
부정맥 분류: MFCC 및 Softmax 특성의
과업 의존적 유효성 평가
이병훈의 석사 학위논문으로 인준함

심사위원장 _____ 구 성 욱

심사위원 _____ 장 원 석

심사위원 _____ 장 승 진

연세대학교 대학원
융합의학과

2025년 07월

차 례

표 차례	iv
그림 차례	v
국문 요약	vi
제1장 서론	1
1.1. 연구 배경 및 필요성	1
1.2. 연구 목표	3
1.3. 연구 기여 방향	5
제2장 분석 방법의 이론적 배경	7
2.1. 연구 내용	7
2.2. ECG 신호	8
2.3. MFCC	9
2.3.1. MFCC 연산	9
2.3.2. ECG 신호에서 MFCC 활용 가능성	11
2.4. Deep Learning Hybrid Model	15
2.4.1. Convolutional LSTM	15
2.4.2. BiLSTM	17
2.5. Activation Function	21
2.5.1. ReLU	21
2.5.2. Softmax	24
2.6. 성능평가지표	27
2.6.1. Accuracy	27
2.6.2. F1-score	28
제3장 실험 방법론	29
3.1. Dataset	29
3.2. MFCC	32

3.3. 1차 실험: 이진 분류	34
3.3.1. convLSTM	34
3.3.2. CNN+Bidirectional LSTM	36
3.4. 2차 실험 방법론	40
3.4.1. CNN+BiLSTM 기반의 다중 분류 성능 평가	40
3.5. 3차 실험 방법론	42
3.5.1. Mel Filter Bank 수 변화에 따른 성능 비교(이진 분류)	42
3.6. Cross Validation	44
3.7. 성능 평가	46
제4장 실험 결과	48
4.1. 1차 실험: 이진 분류	49
4.1.1. MFCC 기반의 convLSTM	49
4.1.2. MFCC 기반의 CNN+BiLSTM hybrid model	52
4.1.3. convLSTM	55
4.1.4. CNN+BiLSTM	58
4.2. 2차 실험: 다중 분류	62
4.2.1. 실험 결과	62
4.3. 3차 실험: Mel Filter Bank 수 변화	68
4.3.1. 실험 결과	68
제5장 고찰	74
5.1. 1차 실험: 이진 분류	74
5.1.1. 세그먼트 길이의 영향	74
5.1.2. 활성화 함수의 영향	75
5.1.3. MFCC + Softmax 기반 하이브리드 구조의 의의	76
5.2. 2차 실험: 다중 분류	77

5.2.1. MFCC + Softmax 구조의 임상적 효과	77
5.3. 3차 실험: Mel Filter Bank 수 변화	78
5.3.1. Mel Filter Bank 감소에 따른 효과	78
5.4. 연구 결론과 기여	79
참고문헌	85
Abstract	87

표 차례

<표 1. MFCC 기반의 ConvLSTM 모델 결과표>	51
<표 2. MFCC 기반의 CNN+BiLSTM 모델 결과표>	54
<표 3. ConvLSTM 모델 결과표>	57
<표 4. CNN+BiLSTM 모델 결과표>	60
<표 5. 최종 결과 종합표>	61
<표 6. Filter Bank = 40 결과>	64
<표 7. Filter Bank = 64 결과>	65
<표 8. Filter Bank = 128 결과>	66
<표 9. Filter Bank = 256 결과>	67
<표 10. CNN+BiLSTM 결과>	70
<표 11. MFCC+CNN+BiLSTM>	71
<표 12. CNN+BiLSTM+softmax>	72
<표 13. MFCC + CNN+BiLSTM+softmax>	73

그림 차례

<그림 1> PQRST 파형의 ECG 신호	9
<그림 2> LSTM 그래프	20
<그림 3> BiLSTM 그래프	20
<그림 4> ReLU 그래프	23
<그림 5> Softmax 그래프	26
<그림 6> MIT-BIH 신호 일부	31
<그림 7> Filter Bank = 40 PR-Curve	64
<그림 8> Filter Bank = 64 PR-Curve>	65
<그림 9> Filter Bank = 128 PR-Curve>	66
<그림 10> Filter Bank = 256 PR-Curve>	67
<그림 11> CNN+BiLSTM PR-Curve>	70
<그림 12> MFCC+CNN+BiLSTM PR-Curve>	71
<그림 13> CNN+BiLSTM+softmax PR-Curve>	72
<그림 14> MFCC + CNN+BiLSTM+softmax PR-Curve	73

국 문 요 약

CNN-BiLSTM 하이브리드 모델을 활용한 심전도 부정맥 분류: MFCC 및 Softmax 특성의 과업 의존적 유효성 평가

본 연구는 ECG(Electrocardiogram, 심전도) 신호 분석에 있어 기존 딥러닝 기반 접근법의 한계를 보완하기 위해, 음성 신호 처리에서 널리 사용되는 MFCC(Mel-Frequency Cepstral Coefficients)를 ECG에 적용하는 시도를 제안했다. 기존의 FFT나 웨이블릿 변환 기반 방식은 시간 정보 손실 혹은 다양한 파형 패턴을 포괄하기 어려운 한계를 지닌 반면, MFCC는 비선형 주파수 특성과 시간 정보를 동시에 반영할 수 있는 장점을 가진다. 본 연구는 MFCC 기반 입력을 활용한 하이브리드 딥러닝 모델(ConvLSTM, CNN+BiLSTM) 구조를 설계하였으며, MFCC 적용 여부 및 세그먼트 길이(2초/3초), 합성곱 계층의 활성화 함수(ReLU/Softmax) 등 다양한 조건 하에서 성능 비교를 수행하였다. 특히 CNN 계층 합성곱 계층에 일반적으로 사용되지 않는 Softmax 함수를 실험적으로 적용한 결과, 주요 특징의 선택적 강조가 가능해져 성능 향상에 기여함을 확인하였다. 실험 결과, MFCC와 Softmax를 함께 적용한 CNN+BiLSTM 모델이 F1-score 기준 최고 성능(0.93078)을 기록하였고, MFCC 미적용 모델과의 비교에서도 일관된 우위를 보였다.

추가적으로 MFCC 필터뱅크 수를 조절한 실험과 다중 클래스 분류 확장 실험을 통해 입력 구성과 과업 조건 변화에 따른 성능 민감도를 검토하였고, Q파형을 제외한 임상적 유의 클래스만을 대상으로 평가함으로써 실제 적용 가능성도 함께 탐색하였다. 이러한 분석은 MFCC 기반 심전도 분류가 단순한 구조적 개선을 넘어, 임상적 판단과 실시간 모니터링 시스템 구현에 기여할 수 있음을 시사한다.

핵심되는 말: ECG, MFCC, 다중 분류, Filter Bank, CNN, BiLSTM, Softmax

1. 서론

1.1. 연구 배경 및 필요성

심혈관 질환(CardioVascular Disease)은 세계 주요 사망 원인 중 하나이며, 암 다음으로 가장 높은 사망자를 이루고 있다. 해당 질환으로 인해 매년 약 1790만 명이 사망하며 이는 전체 질병 사망자의 30%에 해당한다. 또한, 전체 성인 인구의 1~2%에서 발생하며, 고령층에서는 이 비율이 급격히 증가한다. CVD의 한 범주인 부정맥(Arrhythmia)은 심장의 전기적 신호의 이상으로 인해 발생하며, 가벼운 증상에서부터 심각한 심정지에 이르기까지 다양한 형태로 나타난다. 부정맥은 여러 질환 중에서도 뇌졸중, 심부전, 심정지에 해당하는 질환의 초기 증상으로 나타나기도 하며, 다른 CVD와 밀접한 연관이 있어 특히 중요하다. 하지만 부정맥의 복잡한 패턴은 의사의 진단을 어렵게 하고, 환자가 병원을 방문하는 시점에 정상적인 심전도(ECG)를 보일 가능성이 있어 조기 발견이 어렵다. 이러한 이유로 ECG 분석 및 예측에 대한 필요성이 높아지고 있으며, 꾸준히 발전하고 있다.

다양한 연구 방법 중, 최근 인공지능 기술의 비약적인 발전으로 딥러닝과 머신러닝을 통한 ECG 신호 분석 기술이 발전했지만, 여전히 전통적인 신호처리와 인공지능을 활용한 방법은 높은 정확도를 달성하지 못하는 것으로 확인했다. 그 이유는 기존의 연구에서 제안하는 모델은 추가적인 데이터처리 작업을 하지 않고, 대부분의 기본 모델을 활용하는 것에 집중하고 있는 형태의 제안이기 때문이다 [1-2]. 다양한 방식의 모델 활용은 많이 이루어졌지만, 데이터를 활용하여 특징점을 더 잘 가공하여 기존의 방식보다 더 다양한 방식의 접근법을 제안할 필요성을 확인했다.

전통적인 방식의 대표적인 예로 기존의 신호처리 기법인 FFT(Fast Fourier Transform)나 웨이블릿 변환 기반의 특징 추출 방법은 부정맥의 복잡한 시계열 패턴을 완벽하게 반영하지 못한다. FFT의 경우 신호를 주파수 도메인으로 바꾼다는 장점을 가지지만, 심혈관 질환에서 가장 중요한 시간 축에 대한 정보를 소실하게 되면서 이를 반영하지 못한다는 것은 중요한 특징을 반영할 수 없는 단점을 가진다. 웨이블릿 변환 기반의 방법의 경우 시간에 대한 정보를 가지지만 부정맥의 패턴이 다양한 만큼 웨이블릿 함수로 모든 부정맥의 패턴을 함수화하는 것은 불가능에 가까운 일이기 때문에 적합하지 않다.

반면, 본 연구에서 활용하고자 하는 MFCC(Mel Frequency Cepstral Coefficients)의 경우 비선형 주파수 스케일을 적용하여 ECG가 가지는 비선형적 주파수 특성을 반영하고, 시간 정보를 가질 수 있으며, 기존의 원본에 대비하여 특징을 가진 추가 정보를 주기 때문에 더 의미있는 cost를 주는 역할을 할 수 있을 것으로 판단했다.

이뿐만 아니라 최근 발전하고 있는 딥러닝의 CNN과 RNN을 활용한 접근 방식 또한 신호의 특징을 추출하는 과정에서 여전히 개선의 여지가 남아있는 것으로 판단했다 [3]. CNN의 경우, 공간적 데이터를 활용하는 것에 장점이 있으며, LSTM의 경우 시간적인 데이터를 활용하는 것에 장점이 있기 때문에 이것을 모두 더해 새로운 제안을 해보고자 한다.

따라서 본 연구는 기존 딥러닝에는 활용하지 않은 전통적인 방법의 신호처리 방식의 특징 추출과 딥러닝을 합한 방식을 제안하고, 과거의 다양한 모델과의 비교를 통해 얼마나 효과적인지 확인하고자 한다. 이는 MFCC를 특징 추출을 활용한 정보와 원본 데이터를 모두 활용하여 CNN과 RNN의 모델을 결합한 hybrid model을 제안할 때, 기존의 hybrid 모델과 비교하여 얼마나 효과적인지를 확인하고, 더 다양한 방식의 ECG 신호에 대한 접근과 특징 추출이 가능한지 파악하는 것을 연구 목표로 한다.

1.2. 연구 목표

본 연구는 ECG 신호를 이용하여 CVD를 분류할 때, 전통적인 검증된 음성 처리 방식 중 하나인 MFCC를 활용한 정보를 딥러닝을 통한 정보에 추가로 제공할 때, 효과적인 분석과 분류 성능 향상에 기여할 수 있을지를 확인하는 것을 목표로 한다. 최근 ECG 분류에 대한 연구는 ECG 신호가 주는 파형의 물리적인 특징이나 형태학적 정보를 활용하여 Machine Learning 혹은 Deep Learning 모델에 적용하는 방법을 주로 활용한다. 하지만 MFCC와 같은 음성 특징 추출 기법은 잘 시도되지 않으며, 제안된 방법들은 비교적 제한된 환경에서 활용된 사례가 빈번하다 [4]. 따라서 본 연구는 음성 특징 추출 기법인 MFCC를 활용하여 음성 특징 추출 기법에 대한 정보가 ECG 분석 및 연구에서 얼마나 유효할 수 있는가를 기존의 방식과의 비교를 통해 평가하고, MFCC가 ECG 신호에서 유의미한 특징을 제공할 수 있는지의 여부를 확인하는 것에 있다.

상기한 바와 같이 설득력 있는 연구를 위한 여러 단계에 걸친 세부 목표들을 설정하였고, 크게 3가지 세부 목표를 가지고 진행했다.

첫 번째 세부 목표는 “MFCC 음성 특징 추출 기법이 ECG 신호에서 의미 있는 특징 벡터를 실제로 만들어 낼 수 있는가”에 대한 실증적 검증이다. 음성 처리 분야에서 MFCC는 음성 신호의 스펙트럼 구조를 간결하게 요약해주는 강력한 도구로써 많이 활용되지만, ECG 신호의 시간-주파수 특성이 음성 신호와는 상당히 다르다는 점이 문제로 지적될 수 있다. 따라서 본 연구에서 원본 ECG 신호를 주파수 영역으로 변환하고 Mel 스케일을 적용하는 과정을 통해, MFCC가 부정맥 검출에 유의미한 정보를 담고 있는지 확인한다. 이를 위해 MFCC와 원본 ECG 신호를 같이 입력하는 모델과 원본 ECG 신호만을 입

력하는 모델을 각각 분석했다.

두 번째 세부 목표는 “MFCC를 활용한 것과 기존 방식의 분류 성능 차이가 통계적으로 유의미한가”를 평가하는 것이다. 이를 위해 본 연구에서는 “모델 아키텍처”에서 다중 딥러닝 구조인 hybrid model를 구현하여 모델별 성능을 측정하고 비교 분석한다. 예를 들어, 본 연구에서 진행되는 MFCC를 활용한 ConvLSTM 모델의 경우, CNN 연산을 통해 지역, 공간적 특성을 추출하고 LSTM을 이용하여 순방향 시간적 특성을 추출한다. 추가로, CNN과 BiLSTM의 모델을 동시에 활용한 경우, 순방향과 역방향 시간적 특성을 추출한 모델을 활용하여, ECG 신호의 전 후방 시점에서 모두 나타나는 패턴을 학습할 수 있다. 이 앞선 두 가지 모델들에서 모두 MFCC를 활용했지만 이와 비교 모델로서 MFCC를 활용하지 않고 분석을 진행함으로써 두 번째 세부 목표를 확인한다.

마지막 세 번째 세부 목표는 “MFCC 기반의 ECG 분석의 임상적 활용 가능성”을 확인하는 것이다. 이를 위해 본 연구에서 활용하는 데이터셋은 Cross-validation을 활용하여 Overfitting 문제를 최소화하고자 노력했다. 이뿐만 아니라 비정상과 정상의 비율이 상이한 데이터의 특성을 고려하여 F1-score를 활용한 비교분석을 중심으로 활용하고, 임상 환경을 고려하여 민감도(sensitivity), 특이도(specificity), 정밀도(precision) 같은 추가적인 평가 지표와 정확도(Accuracy) 비교 여부를 검토한다. 이와 같은 다양한 평가 지표를 통해 임상적 활용의 유의미한 근거를 제공하고, 일관되고 안정된 성능을 확인할 수 있다.

더 나아가 본 연구는 실제 구현 환경에서의 사용 가능성을 파악하기 위한 모델 구조의 복잡도와 성능 향상의 연관성을 분석한다. MFCC 추출 과정과 CNN과 LSTM을 결합한 구조는 상대적으로 연산량이 많고 구현 복잡도가 높다. 하지만 실제 적용을 위해 병원에서는 실시간 분석 결과를 제공하는 모델

의 정확도와 계산 효율 간의 균형이 중요하다[5]. 앞서 언급한 바와 같이 현실적인 측면에서의 본 연구는 MFCC를 활용한 추가계산 부담이 존재하더라도, 높은 분류성능과 신뢰도를 제공할 수 있다면 임상적인 활용 가능성은 충분하다는 가설을 전제하고 있다. 따라서 높은 수준의 구현 복잡도와 연산량에도 불구하고 높은 정확도를 제공할 만큼의 유효함을 판단할 수 있도록 다양한 근거를 제시한다.

상기 3가지의 세부 목표의 기준을 가지고 “생체 신호 분석의 음성처리 기법을 융합”하고자 하는 궁극적인 목표를 달성하고자 한다.

1.3. 연구 기여 방향

본 연구는 음성처리 방식인 MFCC 방식이 ECG 신호 분석 중 특히 부정맥 감지에서 얼마나 유용하게 사용될 수 있을지에 대해 평가한다. 기존 ECG 분석과 부정맥 탐지 및 분류에 관한 연구는 원본 데이터를 직접 딥러닝에 활용하거나 시간 영역이나 주파수 영역에서의 특징을 추출하는 경향을 보였다. 하지만 본 연구에서는 일반적인 음성 신호 처리에서 사용되는 MFCC를 ECG 분석에 도입하고, 해당 방식이 부정맥 분류에 있어 얼마나 효과적일 수 있는지 분석하고 평가하고자 한다. 이를 통해 본 연구는 다음의 기여를 제공한다.

첫 번째, MFCC를 활용한 ECG 데이터 분석의 데이터 확장 가능성을 제시한다. 기존 연구들은 ECG 신호의 특징을 추출하는 것에 있어, 대부분 전통적인 주파수를 활용한 방식인 FFT, 웨이블릿 변환 방식을 채택해 왔다. 하지만 MFCC를 활용한 ECG 데이터 분석에 관한 연구는 아직 충분히 진행되지 않았다. 최근 연구에서는 MFCC 기반 특성 추출이 기존의 웨이블릿 변환(DWT) 또는 FFT 기반 방법보다 ECG 분류성능에서 우수함을 보인다고 보고하고 있

다 [6]. 이를 위해 본 연구에서는 MFCC를 적용한 경우와 적용하지 않은 원본 데이터를 활용하여, MFCC가 ECG 신호의 패턴을 효과적으로 파악할 수 있는지에 대한 실증적인 근거를 제시한다. 이는 앞으로의 ECG 신호 활용에 있어서 다양한 방식으로 접근할 수 있는 확장 근거에 기여하고자 한다.

두 번째, 시각화 가능한 모델 활용을 통한 해석 가능성이 증가한다. 일반적인 딥러닝의 과정은 최종 모델의 결정 과정이 일반적인 사람의 시각에서 이해할 수 없는 과정이 들어간다. 하지만 본 연구에서 MFCC를 특징으로 활용함으로써 신호의 시간-주파수 정보를 더 명확하게 처리할 수 있는 여지를 제공한다. 특히 앞서 말한 신경망 모델의 큰 단점으로 꼽을 수 있는 결과 이해의 인과관계를 보다 명확한 단계를 하나 더 제공함으로써 전문 의료인들에게 분류 결과의 이해도를 조금 더 높일 수 있다. 이런 가능성은 추후 ECG 신호를 활용한 딥러닝 모델을 만들 때, 추가적인 인과관계 정보를 제공할 여지가 있다.

최종적으로 본 연구는 MFCC를 활용하여 ECG 신호 분석에 적용하는 초기 연구로, 추후 다양한 확장 연구의 기초가 될 것으로 기대한다. 본 연구는 convLSTM, CNN과 BiLSTM에 대한 해석으로 마무리되지만 추후 다양한 딥러닝과 머신러닝 모델에 활용되는 후속 연구로 진행될 수 있다. 또한, 더 나아가 ECG 신호 분석을 통한 부정맥 탐지의 다양한 방법을 제시하는 것에 기여하여, 심장 질환 조기 탐지 및 의료 기술 향상에 실질적인 도움을 줄 수 있을 것으로 기대한다.

2. 분석 방법의 이론적 배경

2.1. 연구 내용

심혈관 질환의 예측과 진단은 꾸준히 증가하고 있다. 이에 따라 여러 합병증과 심혈관 질환의 전조 증상으로 나타나는 부정맥의 진단은 더욱더 중요해지고 있다. 특히 딥러닝의 발달로 다양한 모델을 활용한 연구는 제안되고 있지만 더 높은 정확도를 올리는 방법에 관한 연구는 필요한 실정이다. 이를 위해 단순히 새로운 모델을 활용하는 것이 아닌 데이터의 특징을 추출해 딥러닝에 접근함으로써 접근하려는 새로운 시도는 ECG 신호 분석의 새로운 방향을 제시하기 위해 반드시 필요하다.

본 장에서는 연구 내용의 개념을 종합적으로 이해하기 위해 중점을 둔다. MFCC 특징 추출 방법을 ECG 신호에 사용하여 CNN과 LSTM 모델을 융합한 hybrid 모델과 CNN과 BiLSTM 모델을 융합한 Hybrid 모델 2가지를 활용하여 MFCC 특징 추출하지 않았을 때와 추출했을 때의 차이를 비교하여 그 효용성을 확인하고자 다음과 같은 연구를 진행했다.

- 1) ECG 신호의 특징 분석
- 2) MFCC 처리와 원본 데이터의 동일한 분석을 위한 Segment second 설정
- 3) segment second 와 activation function에 따른 hybrid 모델 아키텍처
- 4) 성능평가 Accuracy, f1-score, precision, recall(sensitivity), specificity 함수 설정 및 확인
- 5) cross validation 횟수 설정 및 모델 및 성능평가 반영
- 6) 결과 확인 및 특징 분석

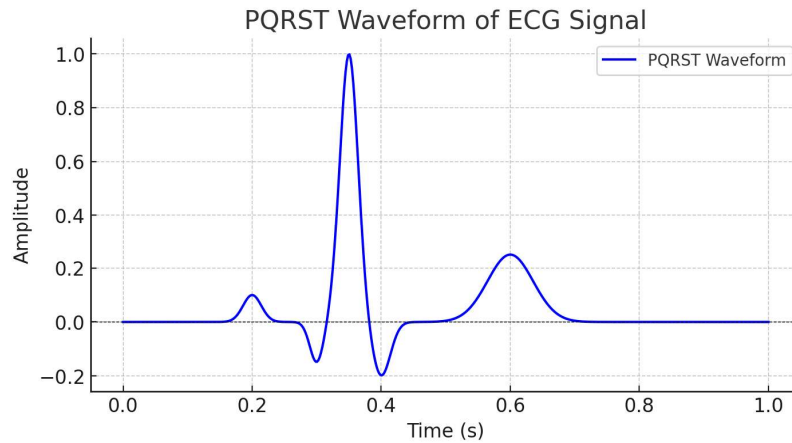
2.2. ECG 신호

심전도(Electrocardiogram, ECG)는 심장의 리듬과 전기적 활동을 기록하는 신호로, 심장의 생리적 상태를 평가하는 데 중요한 역할을 한다. 심장에 대한 많은 정보를 담고 있으며 심장의 평가에 다방면으로 활용된다. 또한, 여러 심혈관 질환을 분석하고 판단할 수 있게 하는 일반적인 비침습적 수단이다.

ECG는 <그림 1>의 형태를 가지고 있다. 이것을 PQRST 파형이라고 한다. 본 파형은 크게 P 파, QRS 복합체, T 파로 구성된다. 각각의 파형에서 P 파형은 심방의 탈분극, QRS 복합체는 심실의 탈분극, T 파형은 심실의 재분극을 의미한다. 이러한 신호를 수학적으로 표현하면, ECG 신호는 시간 t 에 따라 변화하는 연속적인 함수로 기술될 수 있으며, 기본적으로 다음과 같은 합성 수식으로 나타낼 수 있다.

$$ECG(t) = P(t) + QRS(t) + T(t) + N(t)$$

여기서 P , QRS , T 는 각각 P, QRS, T 파형의 시간 함수이며, N 은 노이즈 성분을 나타낸다. 이러한 신호의 분석을 통해 다양한 심장 질환을 감지할 수 있으며, 특히 부정맥 진단에서 ECG 신호의 패턴 분석이 중요한 역할을 한다. 본 연구에서 위의 파형별 특징이 사라지지 않도록 분석하기 위한 새로운 방법으로 MFCC 방법을 도입했다.



<그림 1> PQRST 파형의 ECG 신호

2.3. MFCC

MFCC(Mel-Frequency Cepstral Coefficients)는 신호 처리에서 널리 사용되는 특징 추출 방법으로, 주로 음성 신호 분석에 활용되지만, 최근에는 다양한 생체 신호 분석에서도 적용되고 있다 [7]. MFCC는 멜 주파수 스케일을 기반으로 신호의 스펙트럼 정보를 효과적으로 압축하고, 신호의 주요 특징을 반영하는 계수를 추출하는 방식이다. 본 연구의 독창성의 핵심이 되는 부분이며 해당 부분에 대한 이해를 돕는다.

2.3.1 MFCC 연산

본 연구에서의 MFCC 연산을 이해하기 위해 다음의 연산에 대한 이해를 필요로 한다 [8].

- 1) Framing : 신호를 일정한 길이로 나누어 개별적 분석을 가능하도록 한

다. 일반적으로 프레임을 segment second라고 부르며 본 연구에서는 2초와 3초로 설정해 그 정확성을 확인했다.

2) Windowing : 각 프레임에 Hamming Window를 설정하여 신호의 경계들에서 발생하는 불연속성을 방지한다. 각 프레임별 연산이 들어가면 누락되는 정보로 인해 앞과 뒤의 연결이 끊어지는 것을 막아주는 연산이다.

3) Fast Fourier Transform : 시간 영역의 정보를 주파수 영역의 정보로 변환하는 푸리에 변환이며, 이를 통해 주파수 정보를 분석한다. 연산은 다음과 같다.

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi kn/N}, \quad k = 0, 1, \dots, N-1$$

4) Mel Filter Bank Processing : Mel scale을 기반으로 필터 뱅크를 적용하여 비선형적으로 주파수 정보를 분석한다. 연산은 다음과 같다.

$$E_m = \sum_{k=k_{m-1}}^{k_{m+1}} |X(k)|^2 \cdot H_m(k), \quad m = 1, 2, \dots, M$$

5) Logarithm : 특징 분석의 핵심으로 들어가는 부분으로 사람이 인지할 수 있는 범위에서 높은 진폭의 성분을 줄이고 상대적 차이를 강조하는 연산이다. 연산은 다음과 같다.

$$\log(E_m) = \log\left(\sum_{k=k_{m-1}}^{k_{m+1}} |X(k)|^2 \cdot H_m(k)\right)$$

6) Discrete Cosine Transform : 5번 단계의 로그 변환된 값을 이산 코사인 변환을 수행하여 최종적으로 MFCC 계수를 도출해낸다. 연산은 다음과 같다.

$$c_n = \sum_{m=1}^M \log(E_m) \cdot \cos\left[\frac{\pi n}{M}(m - 0.5)\right], \quad n = 0, 1, \dots, L-1$$

2.3.2 ECG 신호에서 MFCC 활용 가능성

전통적으로 MFCC는 인간 청각계의 특성을 모델링하여 음성의 스펙트럼 정보를 효율적으로 추출하고, 이 과정에서 노이즈에 강인한 특징을 얻을 수 있다는 장점이 있다. 음성 신호는 시간 축에서 짧은 구간별로 나누어 처리되며, 각 구간의 주파수 특성을 멜 스케일로 변환한 뒤 로그 에너지를 적용하고 역 푸리에 변환하는 절차를 거쳐 결국 음향의 중요한 스펙트럼 특징을 수치로 요약한다. 이 과정을 ECG 신호에 접목하면, 심장 활동이 갖는 주파수적 패턴을 반영할 수 있을 것으로 파악했다. ECG 신호 또한 시간에 따라 변동하는 신호이며, 특정 구간별로 심장의 전기적 활동이 주파수 영역에서 뚜렷한 패턴을 만들기 때문이다.

ECG 신호는 앞서 언급한 바와 같이 주로 P파, QRS 복합체, T파 등으로 구분되는 일련의 파형을 통해 심장의 수축과 이완 상태를 전달한다. 하지만 심장 상태나 환자의 개별적 특성, 혹은 측정 환경에 따라 잡음이 섞이거나 왜곡이 생길 수 있다는 점이 문제로 지적되어 왔다. 이때 MFCC 기법이 주파수 영역에서 로그 스펙트럼을 취급해 특징을 추출하는 과정을 거치면, 노이즈가 분산되어 퍼져 나갈 가능성이 커지고, 그 결과 유의미한 주파수 성분이 강조되는 효과가 나타난다. 실제로 음성처리 분야에서 MFCC는 주변 잡음을 어느

정도 무시하면서 발성자의 고유 주파수 성분을 추출하기 위해 설계된 것이기 때문에, ECG 분석에서도 비슷한 원리를 기대할 수 있다. 말하자면, 심장 리듬과 이에 수반되는 전기적 활성을 멜 스케일의 주파수대로 세분화함으로써 일정 수준의 잡음을 걸러내고, 전체 신호를 요약하는 벡터 형태의 특징을 추출하게 된다.

이처럼 MFCC를 ECG 분석에 적용하면, 기존의 ECG 신호를 그대로 사용하는 것보다 몇 가지 이점이 존재한다. 첫째, 차원 축소 효과가 크다. ECG 데이터를 파형 그대로 활용하면 각 샘플링 포인트가 독립적인 차원을 형성하므로, 실시간 모니터링이나 빅데이터 환경에서는 방대한 양의 정보를 그대로 처리해야 하는 부담이 생긴다. 그러나 MFCC를 적용해 주요 주파수 성분만 뽑아내면, ECG 신호의 전체 길이에 비해 상대적으로 적은 수의 특징 벡터로 데이터를 표현할 수 있다. 이는 계산 자원을 절약함과 동시에, 머신러닝 모델이나 신호패턴 분석에서 보다 빠른 처리가 가능하다는 의미다. 예를 들어, 심부전 조기 진단 모델에 MFCC 기반 특징을 활용하면, 무거운 신경망 구조 없이도 민감도 높은 판단을 내릴 수 있는 여지가 커진다.

둘째, 해석 측면에서도 유리하다. MFCC는 음성의 화자 식별이나 단어 인식을 위해 특정 주파수 대역을 중점적으로 반영하는 구조를 갖는다. 이는 ECG 신호에서도 QRS 복합체와 같은 특정 구간의 에너지가 뚜렷하게 드러나는 주파수 대역을 부각시키는 데 도움이 된다. 다시 말해, 원 신호를 시간축에서 직접 확인할 때는 단순한 상승과 하강이 반복되는 것처럼 보이던 ECG가, MFCC 변환 과정을 거치면 초점화된 주파수 대역별 에너지 분포로 나타난다. 이러한 방식의 해석 용이성은, 복잡한 상태 변화를 빠르게 파악해야 하는 의료 환경에서 가치를 높인다.

셋째, 노이즈 억제 능력을 높일 수 있다. 앞서 언급했듯이, ECG 신호는 전극과 피부 상태, 전도율 변화, 외부 전기장 간섭 등 다양한 요인에 의해 잡음

이 쉽게 첨가될 수 있다. MFCC에서는 시간 축에서 미세하게 흔들리는 잡음 보다는, 주파수 축에서 일정 패턴을 형성하는 정보를 강조하기 때문에, 상대적으로 잡음이 묻히거나 희석되는 효과가 나타난다. 특히 로그 스펙트럼 변환을 거치면서 극단적인 진폭값의 차이가 완화되므로, 임상적으로 중요한 작은 스파이크나 파형의 왜곡을 상대적으로 찾기 쉬워질 수 있다. 이를 통해 보다 안정적인 특징을 추출할 수 있고, 이후 분류나 이상 징후 탐지 알고리즘을 적용할 때 오경보 탐지나 미검출 탐지가 줄어드는 결과로 이어질 수 있다.

마지막으로, MFCC 기반 분석은 이미 음성 인식과 오디오 신호 처리 분야에서 널리 사용된 표준적인 방법론을 공유한다는 점에서, 기술적인 도입 장벽이 낮은 편이다. 오픈소스 라이브러리나 표준화된 알고리즘이 충분히 마련되어 있기 때문에, 연구자는 ECG 신호를 소리 데이터처럼 간주하고 이 알고리즘들을 빠르게 시도해볼 수 있다. 머신러닝 혹은 딥러닝 프레임워크에서도 MFCC 추출을 위한 모듈이나 함수를 이미 갖추고 있어, ECG 데이터를 이 형식에 맞춰 전처리한 후 곧바로 모델을 학습시킬 수 있다. 이는 심전도 연구와 음성처리 연구 간의 협업 가능성도 열어주어, 두 분야에서 개발된 다양한 기법이 서로 교차 적용될 수 있는 환경을 만든다.

따라서 ECG 신호에 MFCC를 적용하는 이유는 음성처리 분야에서 이미 효과가 입증된 멜 스케일 기반 주파수 분석 방식을 차용함으로써, ECG가 지닌 복잡적이고 노이즈가 포함된 특성을 보다 구조화하고 해석 가능하게 만들기 위함이다. 기존의 ECG 파형을 직접 사용하는 방식과 비교했을 때, 차원 축소를 통해 계산의 효율을 높이고, 주파수 영역의 의미 있는 특징을 추출하면서 잡음을 완화할 수 있으며, 임상 현장이나 연구에서 중요한 국소적 이상 징후를 보다 쉽게 포착할 수 있다는 장점을 기대할 수 있다. 이러한 특성은 향후 심장질환 조기 진단, 환자 모니터링 시스템, 이동형 의료기기 및 웨어러블 센서 데이터 분석 등 다양한 의료 응용 분야에서 MFCC의 가치와 확장성을 높

이는 근거로 작용할 것이다.

2.4. Deep Learning Hybrid model

Hybrid model은 서로 다른 딥러닝 모델을 2개 이상 연결하여 사용하는 것을 지칭한다. 본 연구에서는 CNN 모델과 RNN의 대표 모델인 LSTM 모델을 사용하고, 기존 LSTM 모델의 단점인 단방향 학습을 보완하기 위한 양방향 학습이 가능한 BiLSTM 모델도 활용하여 MFCC 비교 분석 연구를 진행한다.

2.4.1. Convolutional LSTM(convLSTM)

Convolutional LSTM(convLSTM)은 합성곱 신경망(CNN)과 순환 신경망(RNN)의 핵심 장점을 결합하여 시공간적 패턴을 동시에 학습할 수 있는 구조로 제안되었다 [9]. 전통적인 LSTM(Long Short-Term Memory) 모델은 시계열 데이터를 잘 처리하는 것으로 널리 알려져 있으나, 입력 특징을 벡터 형태로 단순화하여 순차적으로 다루는 한계가 있었다. 반면에 CNN은 공간적 혹은 지역적 패턴을 효과적으로 추출할 수 있으나, 시간 축으로 확장된 데이터를 다룰 때는 추가적인 RNN 계열 기법과의 결합이 필수적이었다. 이런 맥락에서 convLSTM은 CNN의 지역적 특징 추출과 LSTM의 장기 메모리 구조를 융합하여, 시공간적 상관관계가 중요한 데이터를 보다 체계적으로 처리하도록 설계되었다.

기본 아이디어는 LSTM의 게이트(gate) 연산에서 사용되는 완전연결(Fully-Connected) 연산을 대신에, 합성곱(Convolution)으로 대체한다는 데 있다. convLSTM은 입력, 삭제, 출력 3개의 게이트 핵심 요소를 모두 합성곱 기반으로 구현함으로써, 시계열 상의 인접 프레임 또는 시점 정보와 함께, 각 프레임(또는 시점)에 담긴 공간적 특징을 병행 학습한다. 이때 파라미터 수가 늘

어날 우려가 있지만, 시공간 패턴이 복잡하게 얽힌 데이터에서는 이런 방식이 표현력(Representation Power)을 높이는 결과로 이어진다.

본 연구에서 convLSTM을 활용하는 동기는 CNN과 RNN 각각의 강점을 모두 활용하여 ECG 신호 처리의 정확도를 높이기 위함이다. ECG는 시간축으로 펼쳐진 연속 신호이면서도, 특정 구간에서 특징적인 양상이 나타나는 국소적 특성을 지닌다는 점에서 시공간적 접근이 의미가 있다. 일반 LSTM을 사용할 경우, 시간적인 흐름은 잘 포착하더라도 공간 특성에 대한 처리를 일일이 전처리 단계에서 수행해야 할 수 있다. 그러나 convLSTM이라면 CNN이 담당하던 합성곱 연산을 내부 게이트 연산에 도입함으로써, ECG 파형의 특정 segment에 대한 필터링과 장기 의존성(Long-Term Dependency) 파악을 동시에 수행한다. 이는 신호가 길어지거나, 잡음이 포함되었을 때도 안정적으로 핵심 패턴을 유지하게 해주므로, 성능 개선에 직접적인 영향을 미칠 수 있다.

또 다른 장점은 계층적 표현이 가능하다는 것이다. convLSTM 레이어를 여러 단계로 쌓으면, 초기 단계에서는 간단한 지역적 특징(예: 파형의 기울기나 급격한 변화)만을 추출하고, 후속 단계로 갈수록 더 추상적인 패턴(예: 평균 리듬, 이상 신호 출현 빈도 등)을 포착할 수 있다. 이는 단순히 CNN과 LSTM을 직렬로 연결한 하이브리드 모델에 비해, 내부적으로 좀 더 유기적인 통합이 이루어진다는 강점이 있다. 즉, 합성곱 필터와 LSTM의 게이트 구조가 상호보완적으로 작동하여, 각각의 단계에서 시간 축과 공간 축을 반복적으로 고려한다.

하지만 convLSTM을 적용할 때 주의할 점도 존재한다. 첫째, 모델 파라미터 수가 늘어나서 학습 비용이 상당히 높아질 수 있다. 이는 ECG와 같이 대규모 시계열 데이터를 다룰 때 더욱 두드러지므로, 적절한 정규화 기법이나 하이퍼파라미터 튜닝이 필수적이다. 둘째, 지나치게 복잡한 구조를 도입하면 과적합(Overfitting) 위험이 높아지므로, 드롭아웃(Dropout) 혹은 배치 정규화

(Batch Normalization) 등과 같은 방법을 통해 일반화 성능을 높여야 한다. 셋째, 합성곱 커널의 크기나 개수에 따라 ECM 신호의 특징을 놓칠 수도 있으므로, 도메인 지식과 실험적 검증을 병행하여 최적 구성을 찾아야 한다.

결론적으로, convLSTM은 합성곱 기반 특징 추출과 순환 신경망 기반 시간 의존성 학습이라는 두 축을 아우르면서, 복잡한 시공간적 패턴을 다루는 문제에 효과적인 해법이 될 수 있다. 본 연구에서는 ECG 신호의 주파수 스펙트럼 분석(MFCC) 결과물과 convLSTM의 고도화된 시공간 처리 능력을 접목함으로써, 단순한 CNN+RNN 결합보다 한층 정교한 하이브리드 모델을 구현하고자 한다. 이는 의료 데이터 특유의 잡음과 이상 패턴에 대처하기 위한 대안으로서, 추후 다양한 생체 신호 분석이나 헬스케어 분야에도 적용 가능성을 열어줄 수 있을 것으로 기대한다.

2.4.2. BiLSTM

BiLSTM(Bidirectional Long Short-Term Memory)은 전통적인 LSTM 구조를 확장하여 순방향(Forward)과 역방향(Backward) 정보를 동시에 학습할 수 있도록 고안된 모델이다. 일반 LSTM이 과거 시점에서 현재 시점에서의 단방향 흐름을 학습하는 데 비해(<그림 4>), BiLSTM은 현재 시점 주변 맥락을 좀 더 풍부하게 반영한다. 특히, 시계열 데이터가 주어진 방향성 이외에도 역방향 정보를 고려했을 때 해석에 유의미한 단서가 있는 경우, BiLSTM은 단순 LSTM보다 높은 예측력과 표현력을 보이는 것으로 보고되어 왔다 [10].

BiLSTM의 구조적 특징은 <그림 5>와 같이 입력 시퀀스가 두 개의 독립된 LSTM 레이어에 각각 전달되어 하나는 순방향(과거→현재), 다른 하나는 역방향(미래→현재)을 학습한다는 점이다. 이후 이 두 레이어에서 얻은 출력을 결

합(concatenate)하여, 최종 상태 벡터 혹은 중간 단계 벡터로 활용한다. 이를 통해 모델은 과거 시점뿐 아니라 미래 시점의 정보를 통해 현재 상태를 해석함으로써, 시계열 예측이나 분류 문제에서 더욱 견고한 맥락 정보를 구성하게 된다. 예컨대 자연어 처리(NLP)에서는 문장의 앞뒤 단어들을 모두 고려하여 의미를 분석할 수 있고, 음성 신호 처리에서도 특정 구간 앞뒤에 위치한 신호 패턴을 함께 참조함으로써 발음 구간을 명확히 구분하는 데 도움이 된다.

ECG 신호 분석에서 BiLSTM의 잠재적 가치는 명확하다. ECG는 원칙적으로 시간 흐름에 따라 기록되지만, 분석 과정에서 특정 구간(예: R 파의 위치나 P-QRS-T 순서) 주변을 전후 맥락과 연계해 볼 때 의미가 더욱 확연해지는 경우가 많다. 가령, R 파가 높은 진폭으로 찍혔어도 바로 직후 구간에 이상 리듬이 감지되지 않았다면 일시적 잡음으로 간주할 수 있고, 반대로 R 파 전후 구간에서 점진적으로 나타나는 진폭 변화나 리듬 변동은 실제 심장 이상 징후로 이어질 가능성이 높다. BiLSTM은 이러한 앞뒤 맥락의 상호 의존성을 보다 깊이 있게 모델링할 수 있으므로, 단순 LSTM보다 ECG 파형의 세부 정보를 잘 포착할 수 있다.

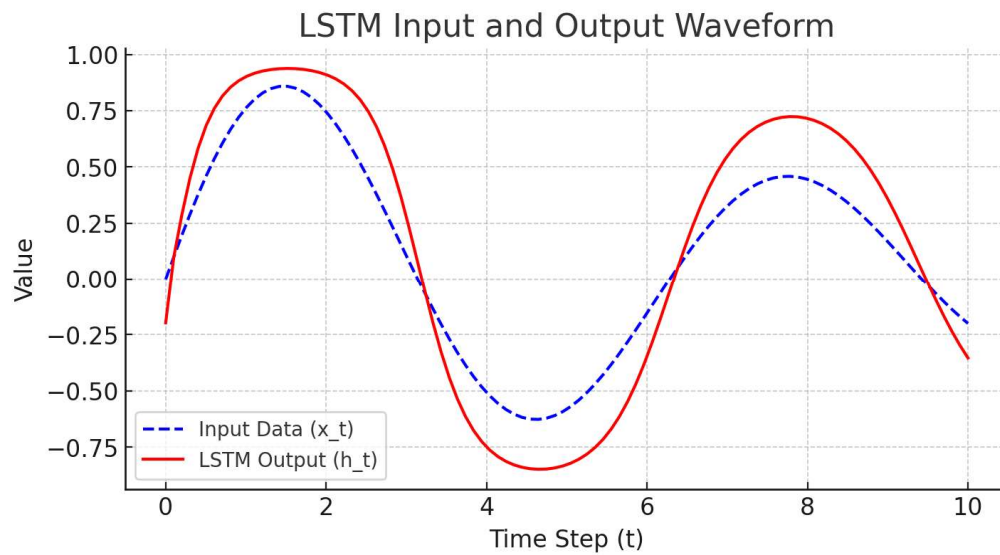
본 연구에서는 BiLSTM을 CNN과 결합하여 하이브리드 모델 구조를 구성한다. CNN이 공간적 패턴 혹은 시공간적 패턴을 추출하는 역할을 맡는다면, BiLSTM은 최종적으로 추출된 시계열 특징 벡터를 양방향에서 해석함으로써 ECG 이상 패턴 판단을 한층 더 정교하게 만들어줄 수 있다. 또한 MFCC와 같은 스펙트럼 기반 특징까지 결합할 경우, 시간뿐 아니라 주파수 영역에서도 양방향 상관관계를 파악할 여지가 생긴다. 이는 특정 대역에서 발생하는 특이 패턴이 이후나 이전 시점의 다른 대역 특성과 결합되어 이상 징후로 나타나는 경우를 놓치지 않게 해줄 수 있다.

추가로, BiLSTM을 활용할 때 고려해야 할 점으로는 추론 비용이 일반 LSTM보다 다소 높을 수 있다는 점을 들 수 있다. 순방향과 역방향 두 개의

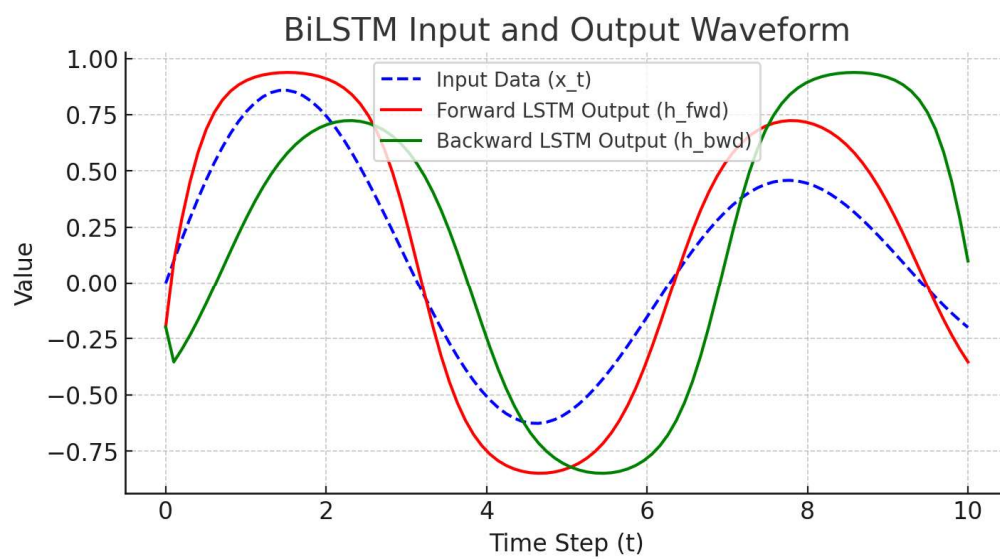
LSTM 레이어를 병렬적으로 학습해야 하므로, 메모리 사용량이 커질 뿐 아니라 계산량도 증가한다. 또한 BiLSTM의 역방향 학습은 미래 시점 정보를 현재 시점 학습에 반영하는 형태이므로, 온라인 혹은 실시간 처리가 필요한 시스템에서는 적용이 까다로울 수 있다. 그러나 ECG 분석의 경우, 이미 기록된 신호를 바탕으로 사후 분석을 수행하는 경우가 많아, 큰 제약이 되지 않을 가능성이 크다. 하지만 본 연구에서는 epoch를 10번으로 제한해 실시간 제약에서 벗어나기 위해 적은 학습 횟수로 실험했다.

종합적으로 BiLSTM은 시계열 맥락을 풍부하게 반영해, ECG에서 생길 수 있는 리듬 이상이나 급격한 변화를 더 정확히 포착하도록 도와준다. 특히, 길이가 긴 ECG 데이터나, 여러 구간에서 비정상 패턴이 산발적으로 나타나는 경우, 양방향 학습은 전후 상황을 동시에 비교하여 잡음과 의미 있는 신호를 구분하는 데 긍정적으로 작용한다. 결국 BiLSTM을 통해 얻게 되는 양방향 정보를 CNN 또는 convLSTM이 추출한 특징 맵과 결합하면, 하이브리드 모델의 다양성과 표현력이 강화된다.

본 연구가 지향하는 바는, BiLSTM의 양방향성으로부터 비롯되는 맥락 해석 능력과 CNN이 제공하는 국소적·시공간적 특성을 유기적으로 엮어내어, ECG 신호 분석에서 발견될 수 있는 복잡한 패턴을 효율적으로 학습하는 것이다. 이 과정에서 MFCC 기반 특징 역시 BiLSTM의 맥락 분석에 결합되어, 주파수 영역의 특정 변화가 전후 맥락 속에서 어떻게 해석되는지를 면밀히 검토할 수 있게 된다. 결과적으로, 이는 단순 모델로는 놓치기 쉬운 미묘한 변동까지 포착하여 심장 이상 여부를 더욱 정확히 추론하는 토대를 마련해 줄 것으로 기대한다.



<그림 2> LSTM 그래프



<그림 3> BiLSTM 그래프

2.5. Activation Function

본 연구에서 활성화 함수는 최종 평가 성능을 높이기 위한 다양한 방식의 시도 중 하나로 활용됐다. 그 중, 합성곱 신경망에서의 통계적 접근을 통한 새로운 활성화 함수 활용에 대한 해석은 ECG 신호 처리에 있어 새로운 해석이 가능하도록 했다. 일반적으로 활용되는 CNN 레이어에서의 ReLU와 sigmoid 활용 대신, ECG 신호의 특징을 더 잘 잡을 수 있도록 하는 Softmax 함수 활용을 통해 새로운 방식의 시도를 해보고자 했다. 이를 위해 다음의 이론을 덧붙이고자 한다.

2.5.1. ReLU

ReLU(Rectified Linear Unit)는 딥러닝 분야에서 가장 광범위하게 쓰이는 활성화 함수 중 하나로서, 간단하면서도 뛰어난 학습 효과를 제공한다. 수식으로는 다음과 같이 정의된다.

$$ReLU(x) = \max(0, x) \quad 2.6$$

이는 입력값 x 가 0 이하일 경우 0을 출력하고, 0보다 큰 양의 값일 경우 해당 값을 그대로 반환하는 방식이다 [8]. 따라서 <그림 2> 같은 구조적 특성인 음수 구간을 완전히 차단함으로써, 역전파 과정에서 발생하던 기울기 소실 (Vanishing Gradient) 문제를 효과적으로 완화한다. 더욱이 ReLU는 미분이 쉽고 계산량이 적어 대규모의 신경망에서 연산 효율성을 높일 수 있다는 장점이 있어, 합성곱 신경망(CNN)을 비롯한 여러 딥러닝 구조에서 표준적으로 활

용되어 왔다.

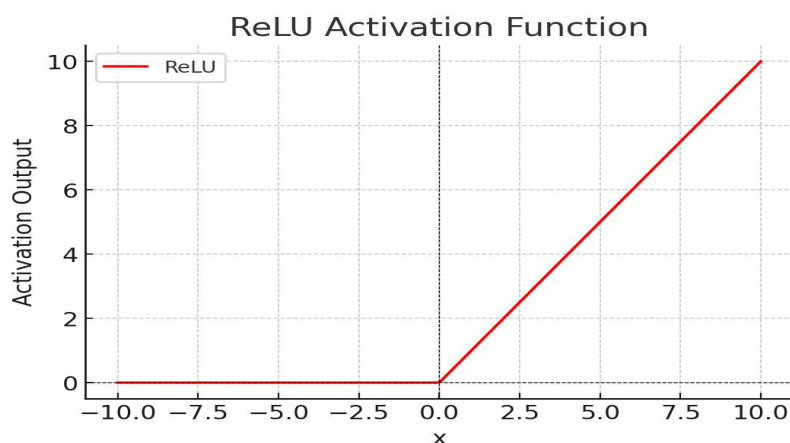
딥러닝 모델에서 ReLU의 효과를 구체적으로 살펴보면, 우선 학습 안정성 측면에서 이점이 있다. 시그모이드(Sigmoid)나 하이퍼볼릭 탄젠트(Tanh) 같은 전통적 활성화 함수는 입력값이 큰 양수이거나 큰 음수가 되면 미분값이 거의 0에 가까워지는 포화 영역(Saturation) 문제가 발생한다. 이러한 포화 영역에서는 역전파가 진행될 때 가중치 업데이트가 미미해져 학습이 정체되는 문제가 생길 수 있다. 그러나 ReLU는 0 이상의 양수 구간에서 기울기가 1로 유지되므로, 비교적 안정적이고 빠른 학습을 가능하게 한다. 음수 구간에서는 출력이 0이 되므로, 필요 없는 신호를 과감히 제거하여 잡음을 줄일 수도 있다. 다만, 이 음수 구간 차단 때문에 학습 과정에서 특정 뉴런이 영구적으로 ‘죽어(dead)’버리는 현상이 발생하기도 하는데, 이를 완화하기 위해 Leaky ReLU, ELU 등의 변형 함수가 제안되기도 했다.

ECG(Electrocardiogram) 신호 분석에서도 ReLU는 자주 활용된다. ECG는 생체 신호 특성상, 여러 요인에 의해 잡음이 섞이거나 이상 징후가 짧은 순간에 발현되는 일이 흔하다. ReLU를 활용하면 급격한 양의 파형은 그대로 반영하여 중요한 특징을 도드라지게 만들고, 음수 신호나 불필요한 잡음이 학습 과정에서 큰 영향을 끼치지 않도록 억제할 수 있다. 이는 필터 학습에서 효율성을 높이고, 모델이 ECG의 특이 패턴(예: 특정 간격의 돌출, 리듬 이상 등)을 놓치지 않게 하는 긍정적인 효과를 낳는다. 또한, ReLU는 단순 구조로 인해 계산 지연이 적어, 대용량 ECG 데이터 세트에도 빠른 처리가 가능한 장점이 있다.

본 연구에서는 ReLU 함수를 통상적인 CNN 아키텍처의 일부 계층에 적용하되, ECG 신호의 특정 패턴을 극대화하기 위한 변칙적 활성화 기법과 병행하여 사용함으로써 기존 접근과의 차별화를 모색하였다. 음수 구간의 정보를 제거해 버리는 ReLU의 단점을 상쇄하기 위해, 자료 전처리 단계에서 일부 스

펙트럼 정보를 미리 추출하거나, 통계 기반 접근을 활용하여 중요한 구간을 놓치지 않는 방향으로 모델 설계를 진행했다. 이러한 설계는 ReLU가 간단하면서도 강력한 활성화 함수라는 점을 유지하면서, ECG라는 도메인 특수성을 최대한 반영하고자 한 시도로 볼 수 있다. 결과적으로, ECG 신호에 대한 분류나 이상 패턴 탐지 등 다양한 실험에서 ReLU가 지닌 계산 효율과 균형 잡힌 학습 안정성을 기반으로 일관성 있는 성능 향상을 관찰할 수 있었다.

딥러닝 모델이 점차 대형화되고, 다양한 생체 신호 분석 상황에 대응해야 하는 환경에서 ReLU는 여전히 높은 우선순위를 가진 활성화 함수다. 비록 최근에는 Swish, Mish 등 더 복잡한 수식에 기반을 둔 함수들도 제안되고 있지만, ReLU의 단순성과 안정성은 경쟁력 있는 대안으로 인정받고 있다. 향후에는 ReLU 기반 모델을 다른 정규화 기법이나 최적화 알고리즘과 조합하여, ECG 신호의 노이즈나 변동성에 더욱 강인한 모델을 설계하는 방향으로 연구가 확장될 수 있다. 이처럼 ReLU는 본 연구에서 핵심 기반으로 삼고 있는 활성화 함수 중 하나로서, 합성곱 연산을 통한 특징 추출 과정과 이상 신호 감지에 중요한 역할을 수행한다.



<그림 4> ReLU 그래프

2.5.2. Softmax

Softmax 함수는 분류 문제에서 출력 노드 간의 경쟁 관계를 명확히 정의하기 위해 고안된 활성화 함수로 널리 알려져 있다. 일반적인 CNN 구조에서는 대개 마지막 분류 계층에서 Softmax를 적용해, 다중 클래스 각각에 대한 확률 분포를 합이 1이 되도록 정규화한다. 이때 특정 클래스에 대한 확률이 높아지면 나머지 클래스에 대한 확률이 상대적으로 낮아지는 구조가 자동으로 형성된다. 이러한 특성은 모델의 출력을 직관적인 확률적 해석으로 연결할 수 있게 만든다. 수식적으로, Softmax 함수는 K 개의 출력을 갖는 모델에서 각 출력 노드 j 에 대해 다음과 같이 정의되며 그래프는 <그림 3>과 같다.

$$Softmax(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

이는 모든 출력 노드에 대한 지수 함수의 합으로 각 노드의 지수 함수값을 나누는 꼴이다. 그 결과, 모든 노드의 출력값은 0과 1 사이의 범위를 갖게 되며, 이들의 합은 항상 1이 된다. 따라서 모델이 예측하는 클래스 라벨이 무엇인지, 각 클래스에 대한 확률이 얼마나 되는지를 명시적으로 파악할 수 있다. 전통적으로는 CNN의 최종 단계에서 Softmax가 적용되지만, 본 연구에서는 합성곱 계층의 중간 단계에서 Softmax를 도입함으로써, 채널별 특징 맵이 경쟁을 통해 더욱 분명한 특성을 띠도록 유도하려는 시도를 했다.

ECG 신호 분석에서 Softmax를 중간 계층에 적용하는 일은 흔치 않은데, 이는 채널 간의 정보가 강제로 정규화되는 과정에서 오히려 학습 성능이 저하

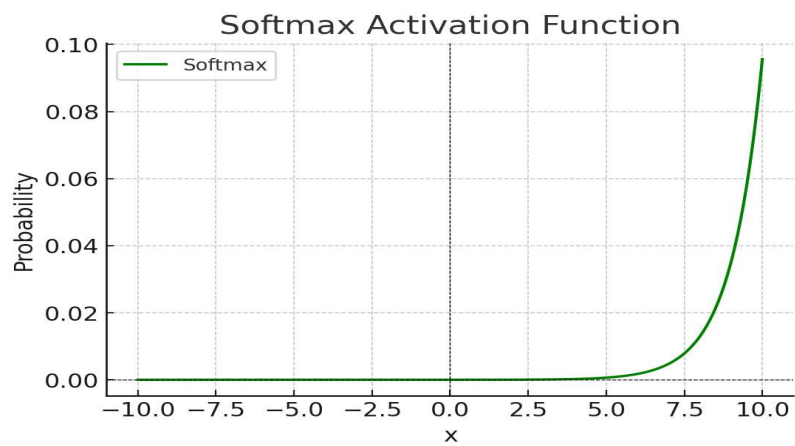
될 것이라는 우려가 있기 때문이다. 하나의 채널이 두드러지게 활성화되면, 다른 채널들의 활성화 정도가 그만큼 희석되는 결과가 나타나, 채널 다각화를 통한 풍부한 특징 추출이 제한될 수 있다는 지적이 대표적이다. 그러나 본 연구에서는 ECG 데이터의 특성, 즉 중요 특징이 비교적 명확한 파형 형태로 나타나고, 여러 채널 간 과도한 중첩이 오히려 잡음이나 혼동을 가져올 수 있다는 점을 고려했다. 이 때문에 채널 간 경쟁이 치열하게 일어날수록, 주파수 혹은 시간 영역에서 돋보이는 신호가 더 잘 부각될 수 있다는 가설을 세웠다.

실제로 본 실험 결과, 중간 합성곱 계층에서 Softmax를 적용한 모델은 특정 채널이 ECG 이상 패턴을 강하게 포착할 경우, 다른 채널에 대한 불필요한 활성화가 억제되어 최종적으로 잡음이 제거된 특징 지도가 형성될 수 있음을 관측했다. 이렇게 얻어진 특징 지도는 이후 레이어의 학습을 더욱 간결하고 효과적으로 만들었다. 예를 들어, 고유의 파형 구조를 강조해야 하는 상황에서 Softmax의 경쟁 메커니즘은 눈에 띄는 특징점을 정확히 포착하거나, 리듬 이상을 빠르게 감지하는 데 도움이 되었다.

물론 이 접근이 항상 우수한 결과를 보장하는 것은 아니며, 채널별 분산이 크게 필요한 다양한 생체 신호나 복잡도가 높은 데이터셋에서는 오히려 활성화 범위를 좁게 만들어 정보 손실이 발생할 위험이 있다. 따라서 Softmax를 중간 계층에 적용할 때는 데이터 특성과 학습 목표를 충분히 고려해야 한다. 본 연구에서는 ECG와 같은 주파수 대역이 제한적이면서도 이상 징후가 특정 구간에 분명하게 나타나는 데이터에 초점을 맞추으로써, Softmax가 가진 경쟁 구조를 효과적으로 활용하는 가능성을 제시했다.

아울러, 최종 레이어에서는 시그모이드(Sigmoid) 함수를 통해 이진 분류 형태로 결과를 산출하도록 설계함으로써, Softmax가 가져오는 채널 간 명확화와 시그모이드의 확률적 단순함이 조화를 이룰 수 있도록 했다. 이처럼 Softmax를 신경망 중간 단계에서 활용하는 창의적인 시도는, ECG 신호 분석에 대한

새로운 각도의 가능성을 열어주었다는 점에서 연구적 의의가 크다. 향후에는 Softmax를 적용하는 시점, 그리고 이를 뒷받침할 만한 정규화 기법이나 보조 손실 함수를 어떻게 조합하느냐에 따라, 생체 신호 분석의 정확도 및 설명 가능성이 더 높아질 수 있을 것으로 기대한다.



<그림 5> Softmax 그래프

2.6. 성능평가지표

본 연구에서 모델의 성능을 평가하기 위한 지표로 Accuracy와 f1-score을 선정했다. 당연히 이를 뒷받침할 증거인 precision과 recall 정보도 같이 수집하였으며 너무 큰 오차나 편차가 날 경우를 대비하여 각 세부 정보들 또한 모두 고려하였고, 단순히 전체 모델을 작업한 것이 아니라 cross validation 작업을 통해 모델 전체의 신뢰도를 높였다.

2.6.1 Accuracy

Accuracy(정확도)는 전체 샘플 중에서 모델이 정답을 맞힌 비율을 의미한다. 수식은 아래와 같다.

$$Accuracy = \frac{\text{정답으로 예측한 샘플 수}}{\text{전체 샘플 수}}$$

Accuracy는 직관적이며 이해하기 쉬운 지표이지만, 클래스 불균형이 심한 문제에서는 높은 Accuracy가 실제로 모델이 우수하다는 것을 보장하지 못한다. 예를 들어, 특정 클래스가 90% 이상을 차지하는 데이터셋에서는 모델이 해당 클래스만을 계속 예측해도 90% 이상의 Accuracy를 달성할 수 있다. 따라서 본 연구에서처럼 클래스 간 데이터 분포가 균등하지 않을 가능성이 있는 음성 및 주파수 분류에서는 Accuracy만을 가지고 모델 성능을 평가하기에 한계가 존재한다. 이를 위해 클래스 간의 분포가 고르지 않은 경우에서도 많이 사용될 수 있는 F1-score도 고려하여 성능평가를 진행했다.

2.6.2 F1-score

F1-Score는 정밀도(Precision)와 재현율(Recall)의 조화를 나타내는 조화평균 값이다. 수식은 다음과 같다.

$$F1 = 2 \times \frac{Precision \times recall}{Precision + recall}$$

Precision은 모델이 특정 클래스라고 예측했을 때 실제로 그 클래스인 경우의 비율을 말하며, Recall은 실제 특정 클래스인 샘플 중에서 모델이 해당 클래스로 예측한 비율을 의미한다. F1-Score가 높다는 것은 모델이 긍정(Positive) 클래스를 예측할 때 ‘얼마나 정확하게 맞는지’(Precision)와 ‘얼마나 놓치지 않고 잘 잡아내는지’(Recall)의 균형이 좋다는 것을 뜻한다.

이러한 특성으로 F1-Score는 클래스 간의 분포가 고르지 않거나, 특정 클래스에서의 오분류가 중요한 문제에서 유용하다 [11]. 본 연구와 같이 부정맥과 정상 맥박의 분류의 분포가 고르지 않기 때문에 시간에 따른 정확한 변화에 따라 분류를 하는 것에 있어 더 정확한 정보를 주는 추가적인 성능평가 지표로서 활용한다. 따라서 특정 클래스의 Recall이 매우 중요해질 수도 있고, Precision 저하가 치명적일 수도 있다. F1-Score는 이러한 Precision과 Recall 간의 균형을 수치화해주기 때문에, Accuracy보다 더 면밀하게 모델 성능을 해석할 수 있다.

3. ECG 파형 전처리 및 분석 과정

본 장에서는 본 연구에서 활용된 구체적인 방법론을 단계별로 살펴본다. 먼저 데이터를 어떻게 수집·전처리하였는지, 그리고 해당 데이터를 실험에서 일관성 있게 다루기 위해 어떠한 기준으로 세분화(segment)했는지를 설명한다. 이후, ECG 신호 특유의 주파수 영역 정보를 보강하기 위해 적용한 MFCC(Mel-Frequency Cepstral Coefficients) 변환 과정을 상세히 다룬다. 마지막으로, 이렇게 준비된 입력 데이터를 활용해 CNN 기반 구조와 LSTM 계열 모델을 하이브리드로 결합한 알고리즘 설계를 소개하고, 각 단계별로 발생할 수 있는 문제점을 어떻게 보완했는지 간략히 언급한다. 본 방법론을 통해 최종적으로 얻고자 하는 목표는, ECG 신호 분석의 정확도와 안정성을 높이는 동시에, 부정맥 등 임상적으로 중요한 이상 패턴을 신뢰도 있게 탐지할 수 있는 실증적 근거를 마련하는 것이다.

3.1. Dataset 전처리

본 연구에서 사용된 ECG 데이터는 MIT-BIH Arrhythmia Database로부터 추출된 것으로, 심장 이상 징후를 포함하여 다양한 리듬 패턴을 폭넓게 다루고 있다 [12]. 이 데이터셋은 임상 환경에서 측정된 실제 ECG 신호를 기반으로 구축되었으며, 여러 환자군의 심전도 특징을 망라하고 있어 심장 리듬 이상을 다각도로 연구하기에 적합하다. 연구의 신뢰도를 높이기 위해 각 레코드(record)를 고르게 선정하였고, 특정 리드(lead)만 사용하여 단일 채널 신호로 통일성을 유지하였다.

데이터 처리는 파이썬 기반 생체 신호 분석 라이브러리인 WFDB를 활용하

여 이루어졌으며, 개별 레코드에 접근할 때는 앞서 언급한 레코드 번호 리스트를 사용하여 필요한 ECG 파형을 전부 불러온다. 파형은 <그림 6>와 같다. 이 과정에서 채널은 보통 MLII(Modified Lead II) 등 가장 대표적인 리드 한 가지만을 선택한다. 이후 레코드별로 얻어진 신호의 총 길이를 확인하고, 이를 일정한 길이(segment) 단위로 나누어 향후 모델이 학습하기 용이하도록 세분화한다. 본 연구에서는 한 세그먼트를 2초(720 샘플), 그리고 3초(1080 샘플) 길이로 나누는 두 가지 전략을 마련하여, 구간 길이에 따른 성능 변화를 비교 분석하였다.

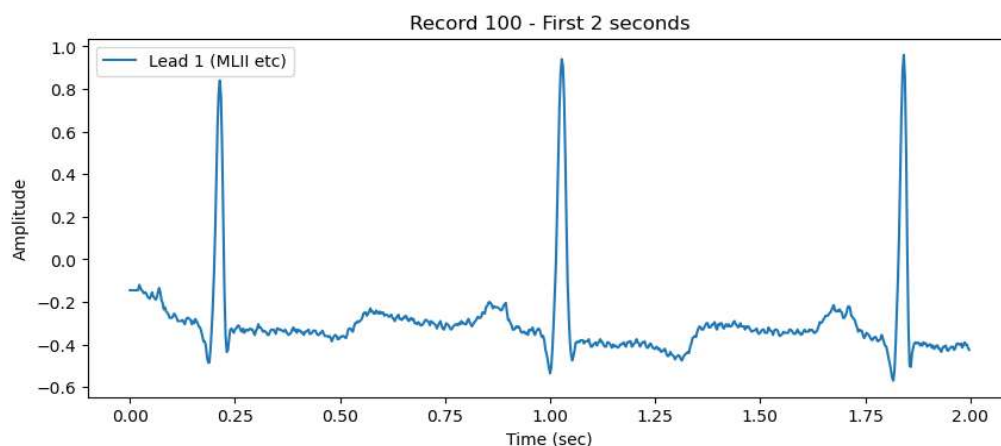
세그먼트 분할 과정에서는 크게 두 가지를 고려했다. 첫째, ECG 신호에서 중요한 파형(예: QRS complex, P파, T파 등)이 1~2초 정도의 짧은 간격 안에 뚜렷하게 나타나는 점을 감안하여, 2초 단위로 쪼개는 실험을 통해 단기 변화를 빠르게 포착할 수 있는지 확인했다. 둘째, 비교적 길이가 긴 3초 단위를 적용했을 때는 리듬 패턴이나 일부 늦게 나타나는 이상 징후를 더 넓은 의미에서 확인할 수 있는 가능성을 염두에 두었다. 이렇게 길이가 상이한 세그먼트를 구성함으로써, 어떤 구간 설정이 실제 분류 정확도나 이상 징후 감지 능력에 영향을 미치는지를 체계적으로 살펴볼 수 있다.

분할된 세그먼트 각각은 중앙 지점에 가까운 어노테이션(Annotation) 정보를 기준으로 레이블이 할당된다. 예컨대 MIT-BIH Arrhythmia Database에는 각 심박마다 부여된 진단 기호가 존재하며, 본 연구에서는 해당 기호가 정상이면 0, 그 외 부정맥 등 이상을 암시하는 기호라면 1로 단순화하여 이진 분류(Binary Classification) 문제로 접근했다. 물론 실제 임상 환경에서는 다양한 종류의 부정맥 라벨이 세분화되어 있으므로, 후속 연구에서는 이 부분을 좀 더 정교하게 다루어 확장할 여지가 있다.

데이터 세트의 불균형 문제도 연구 초기 단계에서 고려되었다. 정상 샘플이 대부분을 차지하거나, 특정 부정맥 라벨이 상대적으로 적은 경우가 있을 수

있기 때문이다. 이를 보완하기 위해 교차 검증(Cross Validation)을 병행하고, 세그먼트 분할로 인해 늘어난 총 샘플 수를 통해 데이터 불균형의 영향을 어느 정도 완화하도록 설계했다. 또한, 분할된 세그먼트가 너무 짧아 중요한 파형을 놓치지 않도록 전처리 단계에서 기본적인 필터링 등을 수행하고, 노이즈가 극도로 심한 일부 구간은 제외함으로써 학습 안정성을 높였다.

이러한 절차를 거쳐 최종적으로 확보된 세그먼트는 입력 데이터 X 와 레이블 y 로 구분된다. X 는 (세그먼트 수 $\times$ 세그먼트 길이) 형태의 넘파이 배열(numpy array)이며, y 는 0과 1로 구성된 이진 라벨 벡터다. 이후 MFCC 추출 과정과 다양한 딥러닝 모델 아키텍처 실험에 활용하기 위해, 이 데이터셋을 파일 형태로 저장해두거나 메모리에 올려둔 뒤 모델에 공급한다. 정리하자면, 본 절에서는 MIT-BIH로부터 ECG 신호를 읽어오고, 2초 및 3초 단위로 나누어 이상 징후 여부를 라벨링하는 전 과정을 망라하였으며, 이는 본 연구의 후속 단계인 MFCC 추출 및 하이브리드 모델 학습의 기반이 된다.



<그림 6> MIT-BIH 신호 일부

3.2. MFCC

ECG 신호는 전형적으로 시간영역(Time Domain)에서 파형의 높낮이나 리듬 변화를 관찰하는 방식으로 해석된다. 그러나 사람의 음성 신호 처리에서 뛰어난 성능을 입증해온 MFCC(Mel-Frequency Cepstral Coefficients) 기법을 ECG에 적용함으로써, 시간축뿐만 아니라 주파수 축(Frequency Domain)에서도 유의미한 정보를 추출할 수 있다. MFCC는 인간의 청각 인지 특성을 반영하여 Mel 스케일로 변환된 주파수 축에서 켈스트럼(Cepstrum)을 계산하는 방식으로, 주파수 대역별 특성을 압축적으로 표현해 준다. 이러한 특징은 ECG처럼 특정 주파수 대역에서 심장 박동과 관련된 정보가 두드러지는 경우에 주목할 만하다.

본 연구에서는, 앞서 2초 및 3초 세그먼트로 분할해둔 ECG 파형을 그대로 MFCC 변환의 입력으로 삼는다.

본 연구의 MFCC 구현은 파이썬의 librosa나 speechpy 같은 음성 처리 라이브러리를 참고하여 MFCC를 계산하되, ECG 신호에 맞춰 일부 파라미터를 조정하였다. 예컨대 일반적인 음성 인식에서는 샘플링 레이트가 8kHz~16kHz인 반면, MIT-BIH 데이터를 기준으로 ECG의 경우 360Hz로 비교적 낮기 때문에, 필터뱅크 설정도 이에 맞춰 줄일 필요가 있었다. 또한, MFCC 추출 단계에서 계산된 프레임별 에너지가 지나치게 낮으면 잡음 처리 과정에서 유실될 위험이 있으므로, 최소 에너지 임계값을 설정하거나, 리프팅(Liftering) 기법을 통해 고차수 계수의 영향을 적절히 보정하는 과정도 고려했다.

이 과정을 통해 최종 산출되는 MFCC 시퀀스는, ECG 세그먼트별 시간-주파수 특성을 압축적으로 함축하고 있다. 단순히 시간축 신호를 직접 입력으로

사용하는 방식에 비해, 주파수 영역에서 놓치기 쉬운 패턴을 더 쉽게 포착할 수 있다는 점이 핵심 장점이다. 특히 부정맥과 같이 특정 간격에서 등장하는 리듬 이상은 Mel 스케일 상에서 일정한 대역폭을 차지할 수 있어, MFCC가 이를 상대적으로 명확히 드러낼 수 있다. 반면 MFCC 변환으로 인해 일부 시간 정보가 압축될 수 있으므로, 그것을 보완하기 위해 ConvLSTM이나 BiLSTM 같은 시계열 처리에 강점이 있는 모델과 결합하는 것이다.

한편, 이렇게 변환된 MFCC 시퀀스는 단순 벡터가 아니라 시계열적 구조를 유지해야 한다는 점이 중요하다. CNN + BiLSTM 조합을 사용하는 경우, CNN이 먼저 2D 혹은 1D 합성곱 연산을 통해 MFCC의 프레임 간 국소적 패턴을 추출한 뒤, 이를 BiLSTM에 넘겨서 전방향 및 역방향 맥락을 동시에 해석한다. 반면 ConvLSTM 모델을 쓸 때는, 합성곱 연산과 LSTM 구조가 한데 섞여 있어, MFCC 프레임(시점)마다 합성곱 필터가 적용되면서 그 결과가 LSTM 게이트를 통해 반복적으로 전달되는 메커니즘을 활용한다. 이때 MFCC로부터 생성된 시퀀스 길이와 각 프레임의 차원 수를 정확히 모델에 맞게 변환해주어야, 학습 과정에서 오류 없이 시계열 정보를 흡수할 수 있다.

본 연구에서는 2초 및 3초 세그먼트 데이터를 MFCC로 변환한 뒤, ConvLSTM과 CNN+BiLSTM 모델의 입력 형식에 부합하도록 시계열 프레임을 배치해 주었다. 이 과정에서 일부 파라미터 최적화와 전처리 기법을 병행함으로써, 모델이 ECG 이상 패턴을 놓치지 않고 파악할 수 있도록 설계하였다. 이렇게 구축된 MFCC 기반 입력은 다음 절에서 다룰 하이브리드 모델의 정확도 검증에서도 중요한 역할을 하며, 특히 부정맥과 정상 맥박 구분이 분명치 않은 애매한 구간에서 더욱 유효하게 작동함을 예비 실험을 통해 확인하였다.

3.3. 1차 실험: 이진 분류

3.3.1. ConvLSTM

본 연구의 목적에 맞게 첫 번째 모델은 시계열 데이터를 2차원 합성곱과 순환 구조를 융합한 ConvLSTM 모델로 실험을 진행하였다. 이 모델은 CNN으로 지역적 패턴을 추출하면서 합성곱 연산을 수행하고, LSTM과 같은 시간축을 따라 연속적인 2차원 특성 맵을 학습하는데 효과적이다. 즉, ConvLSTM은 내부적으로 2D 합성곱 연산을 수행하면서 순환 계층이 작동한다. 이러한 특성으로, 데이터의 시공간적인 특징을 동시에 확인할 수 있는 장점이 있다.

입력단에서의 시계열 형태는 (timesteps, rows, cols, channels) 구조이며, 주어진 ECG 신호를 여러 시점(timesteps)으로 나누고, 한 시점 내에서도 (rows, cols) 형태로 특성을 배치한다. 실제 연구에서 cols=1 형태를 사용하였고, (3,1) 커널 크기를 통해 세로 방향으로 특성을 추출한다. 단계는 다음과 같다.

1) ConvLSTM2D(filters=16, kernel_size=(3,1), activation='softmax')

첫 번째 레이어로 2차원 합성곱과 LSTM을 융합한 ConvLSTM2D를 사용한다. filters=16은 출력 특성맵의 채널 수, kernel_size=(3,1)은 합성곱 필터가 높이 방향으로 3, 폭 방향으로 1의 크기를 가짐을 의미한다. activation은 softmax와 reLU를 모두 적용해보며, ECG 신호 특성에 맞는 활성화 함수를 판단할 수 있도록 했다. return_sequences=False 옵션으로 마지막 시점의 출력만을 취하도록 설정하였다. 이는 최종 시점에서의 은닉 상태가 응집된 형태로 특징을 잘 요약한다고 가정한다.

2) Flatten

ConvLSTM2D 계층의 최종 출력을 2차원 이상으로 가지고 있을 것이므로, 이를 분류를 위해 1차원 벡터로 변환한다. 이를 통해 (batch, flattened_features) 형태가 되어, Dense 레이어에 입력할 수 있도록 한다.

3) Dense(32, activation='relu') 및 Dropout(0.2)

Dense 계층으로 32차원의 은닉 유닛을 구성하고, 활성화 함수로 ReLU를 사용하여 비선형성을 부여한다. Dropout(0.2)을 추가하여 학습 과정 중 20%의 뉴런 출력을 0으로 설정함으로써, 특정 가중치에 과도한 의존을 방지하고 모델의 일반화 성능을 높인다.

4) Dense(1, activation='sigmoid')

본 연구는 부정맥의 여부를 판단하는 이진 문제로 시그모이드 함수를 사용해 0~1 범위의 확률값을 산출함으로써 이진 분류 문제를 해결한다.

ConvLSTM 모델은 시계열 데이터의 각 시점에서 2D 합성곱을 수행하고, 그 결과를 순환 구조로 시퀀스 전반에 걸쳐 전달함으로써 시공간적인 특징을 동시에 학습한다. 특히 합성곱 단계에서 softmax와 ReLU를 활성화 함수로 사용해, 국소 구간에서 가장 두드러지는 특성을 강조하는 방식으로 정보 처리를 수행하면서 그 효과를 확인하고자 한다. 이후 Flatten → Dense → Dropout → 시그모이드 출력을 거치며, ECG 신호 분류를 위한 최종 예측값을 산출한다. 4절에서는 이 ConvLSTM 모델이 정확도(Accuracy), F1-score, ROC-AUC 등 다양한 지표에서 어떤 성능을 보이는지 상세히 평가할 예정이다.

3.3.2. CNN + Bidirectional LSTM

본 모델은 CNN과 BiLSTM을 결합한 하이브리드 모델로 다음과 같은 과정으로 설계했다. 크게 요약하자면 먼저, Conv1D 계층을 시간 축으로 나누어 적용하고, Bidirectional LSTM을 활용해 전체 시퀀스 정보를 학습한 뒤, Dense 계층을 이용해 이진 분류를 수행하도록 했다.

앞서 제시한 ConvLSTM 모델과 동일하게 합성곱 신경망(CNN)을 통해 시간 축을 따라 특성을 추출하고, 순환 신경망(LSTM) 모델로 구성은 했지만 앞서 제시한 단방향인 아닌 양방향 모델로 전체 시퀀스의 전후 맥락을 동시에 학습하는 모델을 설계하였다. 이를 구현하기 위해 Keras의 TimeDistributed 래퍼를 활용해 각 “시간 단위”별로 독립적인 합성곱·풀링 연산을 수행한 뒤, Bidirectional LSTM으로 넘기는 구조를 구성하였다. 코드로 작성된 핵심 로직은 아래와 같으며, 이를 텍스트로 상세히 풀어 설명하면 다음과 같다.

먼저 입력단에서 시계열 형태는 (batch_size, timesteps, features, channel) 구조를 가지며, 이 중에서 timesteps 차원을 따라 시계열 정보가 존재한다. features는 한 시점에서의 특성, channel은 1차원 신호이므로 1로 설정했다. 이후 레이어는 다음의 단계 구조를 가진다.

1) TimeDistributed(Conv1D)

첫 번째 레이어로 TimeDistributed(Conv1D(filters=32, kernel_size=3, activation='softmax'))를 둔다. TimeDistributed 래퍼는 입력으로 들어온 시계열에 대해, 각 시간 스텝별로 동일한 Conv1D 연산을 독립적으로 적용할 수 있도록 한다. Conv1D 계층 자체는 필터 32개, 커널 크기(kernel_size) 3으로 설정되어 있으며, 활성화 함수로는 ReLU나 softmax를 사용한다. softmax는 일반적으로 합성곱 계층에서 자주 쓰이는 ReLU와는 다른 선택이지만, 본 연구의 ECG 신호 특성 상 softmax와의 비교가 유의미할 수 있을 것으로 판단

했고, 비교를 위해 2개의 특성을 바꾸면서 학습했다. 앞서 언급한 바와 같이 (batch, timesteps, features, channel) 형태의 입력을 받아, (batch, timesteps, new_feature_map, filters) 형태의 출력을 생성하게 된다. 다만 TimeDistributed 래퍼로 감싸져 있으므로, 시간 축(timesteps) 차원을 그대로 유지하며 각각의 시점에서 얻어진 합성곱 결과를 순차적으로 쌓아둔다.

2) TimeDistributed(MaxPooling1D)

TimeDistributed 래퍼를 씌운 MaxPooling1D(pool_size=2)를 추가한다. 앞선 레이어인 합성곱 과정에서 생긴 특징맵을 풀링하여 (batch, timesteps, features)의 차원을 축소하고 중요 특성을 남기는 역할을 한다. 시간 축을 따라 각 시점별로 독립적인 풀링 연산이 수행되므로, 합성곱 이후의 특징맵 크기를 줄여 모델의 파라미터 수와 연산량을 합리적으로 관리한다.

3) TimeDistributed(Flatten)

풀링된 2차원 혹은 (특성 × 채널) 형태의 출력을 일렬 벡터로 바꾸는 작업이다. Flatten 레이어 또한 TimeDistributed 안에서 동작하므로, 각 시간 스텝별로 개별 Flatten이 이루어지고, 최종적으로 (batch, timesteps, flattened_features) 형태로 데이터가 변환된다. 이 과정을 통해 이후에 LSTM 계층으로 넘길 때, “시점 × 벡터” 구조로 입력을 손쉽게 전달할 수 있게 된다.

4) Bidirectional(LSTM, return_sequences=False)

전처리가 완료된 시퀀스를 Bidirectional LSTM에 전달한다. 여기서는 LSTM의 hidden units를 64로 설정하였고, return_sequences=False 옵션을 사용한다. return_sequences=False는 LSTM이 모든 시점의 출력을 내보내지 않고, 마지막 시점의 은닉 상태만을 최종적으로 반환하도록 한다. Bidirectional 래퍼는 순방향과 역방향 두 LSTM을 병렬로 학습시켜, 시간 순서의 전후 맥락을 동시에 추론할 수 있도록 한다. 이는 ECG 신호 내 특정 파형이 앞뒤 문맥을 통해 더 정확히 구분되는 경우, 효과적이며 분류의 성능을 끌어올릴 수

있을 것으로 기대한다.

5) Dense(32, activation='relu') 및 Dropout(0.2)

LSTM의 최종 은닉 상태(양방향이므로 내부적으로는 64×2 차원이 통합된 형태)를 입력으로 받는 완전 연결(Dense) 계층이다. 32차원의 은닉 노드를 두어 추가적인 표현 학습을 수행하며, 활성화 함수로는 ReLU를 사용한다. 또한 Dropout(0.2)을 적용해 과적합을 방지하고 모델이 일반화되도록 유도한다. Dropout은 학습 과정에서 무작위로 20%의 뉴런 출력을 0으로 설정하여, 특정 가중치에 지나치게 의존하는 현상을 완화한다.

6) Dense(1, activation='sigmoid')

마지막으로 1차원 출력 뉴런에 시그모이드 활성화를 적용해, 부정맥 여부를 확인하는 이진 분류 확률(0~1)을 산출한다. 다중 클래스인 경우, softmax를 활용하는 것이 일반적이지만, 본 연구에서는 부정맥과 정상인 경우로만 나누어 연구를 진행하였고, 해당 경우에는 이진 분류가 효과적인 결과를 낼 수 있으므로 sigmoid로 최종 산출하였다.

결과적으로 본 모델은 “시계열 신호를 TimeDistributed로 각 시점마다 CNN 전처리 → Bidirectional LSTM으로 전체 시퀀스의 맥락 학습 → 완전연결(Dense) 계층을 거쳐 이진 분류”라는 구조를 취한다. 합성곱 과정에서 활성화 함수로 softmax를 비교 분석하는 점이 일반적인 CNN 설계와는 차이를 보이도록 설계하고 연구했다. 또한 LSTM을 양방향으로 구성함으로써 앞뒤 맥락 모두를 활용할 수 있도록 설계한 것도 본 연구에서의 주요 시도 중 하나이다.

이상의 과정을 통해, 각 시점에서의 국소 패턴을 효과적으로 추출하고(CNN 단계), 시간 순서 양방향 정보를 종합해 ECG 시퀀스 전체의 특징을 학습(Bidirectional LSTM 단계)함으로써 이진 분류 문제를 해결하도록 하였다. 이

후 3.2절과 같이 4절에서 모델을 대상으로 성능평가(Accuracy, F1-score, Precision, Recall 등)를 진행하고, ROC-AUC 곡선도 제시함으로써 모델의 분류 능력을 다각적으로 확인한다.

3.4. 2차 실험 방법론

3.4.1 CNN+BiLSTM 기반의 다중 분류 성능 평가

본 연구의 주된 실험은 이진 분류 문제를 중심으로 진행되었지만, MFCC 기반 하이브리드 모델의 적용 가능성과 확장성을 보다 심층적으로 검토하기 위해 다중 분류 실험을 추가적으로 설계하였다. 이뿐만 아니라 부정맥을 판단하는 임상적인 측면에서 AAMI를 기준으로 한 5가지 클래스 분류를 통해 의미가 있는지의 여부를 판단하고자 했다. 이를 통해 MFCC를 특징 추출 방법으로 활용하는 것이 이진 분류뿐만 아니라, 보다 복잡한 분류 문제에서도 유효할 수 있는지를 평가하고자 한다.

다중 분류 실험은 기존 이진 분류 실험에서 높은 성능을 보였던 CNN+BiLSTM 모델을 그대로 활용하였다. 입력 데이터는 2초 (segment_sec=2) 구간으로 분할된 ECG 신호의 MFCC 변환 결과를 사용하였으며, 입력층은 다중 클래스 분류에서의 CNN 연산층에서의 softmax 성능을 보기 위해 ReLU와 번갈아며 비교 구성하였다.

여기서 클래스 구분은 정상(Normal) 심박과 다양한 종류의 부정맥을 각각 별도의 클래스로 설정하여 진행하였다. 주요 부정맥 유형을 AAMI 기수퍼 클래스 기준으로 분리하여 총 5개의 클래스로 구분하였다.

모델 성능 평가는 F1-score를 주된 지표로 삼았다. 이는 다중 분류 상황에서도 클래스 간 데이터 불균형이 존재할 가능성이 높기 때문에, precision과 recall 간의 균형을 고려할 필요가 있기 때문이다. 이를 위해 Macro-F1 score와 weighted-F1 score를 함께 제시하여 전체적인 분류 성능을 보다 종합적으로 분석하였으며, 부가적으로 정확도(Accuracy)도 함께 확인하여 세부 분류

성능을 검토하였다.

본 다중 분류 실험을 통해, MFCC 기반 CNN+BiLSTM 구조가 이진 분류 문제를 넘어 보다 복잡한 ECG 신호 분류 과제에도 적용 가능한지에 대한 실증적 근거를 제시하고자 하였다. 또한 이를 통해 제안한 특성 변환 방식과 하이브리드 모델 구조의 확장성 및 실용적 유효성을 검증하는 데 목적이 있다.

3.5 3차 실험 방법론

3.4.1 Mel Filter Bank 수 변화에 따른 성능비교(이진 분류)

MFCC는 음성 신호 분석에서 널리 사용되는 특징 추출 기법으로, 신호의 주파수 대역을 멜 스케일(Mel scale)로 변환하여 인간 청각 특성에 가까운 특성 공간을 형성한다 [13]. 이때 변환 과정에서 사용되는 필터뱅크의 수는 MFCC 특성의 해상도와 직결되며, 결과적으로 모델이 학습할 수 있는 정보량에 큰 영향을 미친다. 필터뱅크 수가 작은 경우 주요 주파수 대역만을 포괄하게 되어 상대적으로 단순한 특성이 추출되기 때문에 정보의 누수가 있을 수 있다. 상대적으로 필터뱅크 수를 증가시키면 보다 세밀한 주파수 변화를 반영할 수 있지만, 동시에 노이즈 민감성이 증가하고 과적합(overfitting) 가능성도 높아질 수 있다.

따라서 본 절에서는 앞서 가장 최적이었던 모델을 기반으로 MFCC 필터뱅크 수를 조정하여, ECG 이진 분류 문제에 있어서 최적의 특성 수를 탐색하고자 한다. 실험은 이진 분류 성능이 가장 우수했던 모델 구조인 CNN+BiLSTM 하이브리드 모델을 고정하여 수행하였다. 데이터 전처리 또한 기존 이진 분류 실험과 동일하게 2초(segment_sec=2) 구간으로 설정하였으며, 이를 통해 변수는 필터뱅크 수만 변화시켜 실험의 일관성을 확보하였다.

실험에 사용한 MFCC 필터뱅크 수는 40, 64, 128, 256의 4가지로 설정하였다. 이는 일반적으로 많이 사용되는 수인 128로 기존 모델을 평가 기준으로 삼고, 128보다 큰 수와 작은 수를 넣어 모델을 학습하고 검증하였으며, 클래스 간 데이터 수 불균형이 존재하는 MIT-BIH 데이터셋의 특성을 고려하여 F1-score를 주요 성능 평가 지표로 설정하였다. F1-score를 중심으로 분석하

되, 부가적으로 정확도(Accuracy)를 확인하고 macro average와 weighted average, AUC 등의 지표도 함께 검토하여 전반적인 모델 성능과 특성 변동에 따른 영향을 다각도로 평가하고자 한다.

본 추가 실험은 MFCC 기반 특성 설계가 ECG 신호 분석에서도 주요 파라미터 튜닝 요소가 될 수 있음을 검증하고, 동일한 MFCC 수를 유지한 경우, 더 다양한 해상도의 mel spectrogram 기반에서 특징을 추출했을 때의 효과를 평가하고자 한 것이다.

3.6. Cross validation

본 연구에서 사용한 5-Fold 교차검증(cross-validation)을 각 단계별 절차와 의의에 대해 상세히 기술하고자 한다 [14]. 먼저 데이터의 양과 추후 검증 과정에서 testset 과 trainset의 비율을 8:2로 하는 것이 데이터 불균형과 학습의 양으로 비교했을 때, 적합하다고 판단했다. 따라서 5-fold cross-validation으로 결정했고, 이는 단순 train/test 분할에 비해 훨씬 안정적인 성능 추정이 가능하기 때문이다. 또한 폴드를 반복하여 얻은 여러 지표의 평균값과 표준편차를 살펴봄으로써, 모델이 데이터에 얼마나 일관된 성능을 내는지 가늠할 수 있기 때문에 본 모델이 얼마나 신뢰도 높게 작동할 수 있는지에 대한 근거가 될 수 있을 것으로 판단했다.

먼저, 하나의 폴드마다 고유한 학습용(train)·검증용(val) 세트를 만든다. 이때 같은 데이터가 검증 세트로 쓰이지 않는 폴드는 학습에만 사용되므로, 데이터 낭비 없이 효과적으로 모델을 훈련·평가할 수 있다. 본 연구에서 X 는 (샘플 수 $\times$ 특성 수) 혹은 (샘플 수 $\times$ 시계열 길이 $\times$ 특성) 구조의 배열일 수 있으며, y 는 0 또는 1의 이진 라벨을 가진다(부정맥 여부 등). 각 폴드를 순회 함으로써 데이터 전체가 검증용 세트 역할을 한 번씩 수행하므로, 모형의 불안정성을 줄이고 평균적 성능을 더욱 신뢰도 높게 측정할 수 있다.

폴드별로 분할된 학습·검증 세트가 준비되면, 우선 위의 검증 평가의 편의성을 위해 build_model 함수를 통해 모델 인스턴스를 생성한다. 이 모델은 ConvLSTM, CNN+BiLSTM 등 본 연구에서 제시한 신경망 아키텍처 중 하나를 구현한 함수로, 특정 계층 구조와 하이퍼파라미터(filters, kernel_size, activation, hidden units 등)가 미리 정의되어 있다.

이후 컴파일 과정에서는 옵티마이저는 Adam, 손실 함수는

binary_crossentropy, 그리고 학습 중 모니터링할 기본 지표는 accuracy로 설정한다. 여기서 Accuracy는 학습 과정에서의 지표를 확인하기 위한 용도로 쓰이며, 실제 최종 성능평가는 아래에서 별도로 precision, recall, F1-score 등을 계산하여 수행한다.

또한, 학습 fitting 과정에서는 가볍게 사용하기 위해 epoch를 과도하게 늘리지 않는 것을 핵심 요소로 생각했다. 따라서, batch_size = 32, epochs = 10으로 설정하여 각 폴드별 10번의 연산을 진행하도록 했고, 조기 종료는 하지 않았다. 과한 연산으로 overfitting의 가능성 또한 배제할 수 있도록 했다.

3.7 성능평가

교차 검증 과정에서 같이 진행한 성능평가는 다음의 과정으로 진행했다.

학습이 완료되면, 검증 데이터(X_{val_fold})에 대해 예측을 수행한다. 이때 모델의 출력은 0과 1 사이의 실수 확률값으로 임계값을 0.5로 설정해 0.5보다 큰 경우 1, 그렇지 않으면 0으로 이진 분류했다. 이 과정을 통해 예측 라벨을 생성한다. 예측 라벨을 통해 confusion matrix를 구하고, 실제 라벨 대비 예측 라벨이 얼마나 일치하는지를 행렬 형태로 확인한다. confusion matrix의 각 원소는 다음과 같다.

- 1) TN (True Negative): 실제 음성인데 음성으로 예측
- 2) FP (False Positive): 실제 음성인데 양성으로 예측
- 3) FN (False Negative): 실제 양성인데 음성으로 예측
- 4) TP (True Positive): 실제 양성인데 양성으로 예측

위 원소들을 바탕으로 Accuracy(정확도), Precision(정밀도), Recall(재현율), Specificity(특이도), F1-score, Sensitivity을 계산하여 도출한다. 계산은 다음과 같이 했다.

전체 중 정답을 맞힌 비율인 Accuracy의 수식은 다음과 같다.

$$(TP + TN) / (TP + TN + FP + FN)$$

모델이 양성이라고 예측한 것 중 실제로 양성인 비율인 Precision의 수식은 다음과 같다.

$$TP / (TP + FP)$$

실제 양성 샘플 중에서 양성으로 맞힌 비율 Recall(Sensitivity)은 다음과 같다.

$$TP / (TP + FN)$$

실제 음성 샘플 중에서 모델이 음성으로 맞힌 비율인 Specificity는 다음과 같다.

$$TN / (TN + FP)$$

Precision과 Recall의 조화평균인 F1-score은 위의 2.6.2에서 제공한 수식과 같다.

위와 같이 여러 지표들을 성능평가로 쓰는 이유는 단순히 Accuracy만을 확인하면 클래스 불균형인 경우 성능이 객관적이지 않고, 이진 분류 문제의 성능을 다각적으로 살펴볼 수 없기 때문에 Precision, Recall, F1-score와 같은 여러 지표들과 같이 성능을 확인했다.

4. 실험 결과

본 연구에서는 모델의 성능에 영향을 주는 다양한 실험 조건들을 조합하여 정량적 분석을 수행하였다. 전체 실험은 이진분류, 다중분류, Mel Filter Bank 수 변화에 따른 성능 비교로 총 3가지 큰 실험을 기준으로 진행하였다.

첫 번째 실험의 주요 실험 변수는 크게 네 가지로 분류되며, 이는 (1) 세그먼트 길이 설정(2초 vs 3초), (2) MFCC 적용 여부, (3) 합성곱 신경망의 활성화 함수로 Softmax 사용 여부, (4) 전통적인 ReLU 함수 사용 여부에 해당한다. 이 변수들은 단독으로도 모델 성능에 영향을 줄 수 있으며, 상호작용적으로 작용할 경우 보다 복합적인 성능 차이를 유발할 수 있다. 각 모델은 다음과 같은 조건 조합에 따라 실험되었다. 첫 번째 변수는 세그먼트 길이이며, ECG 데이터를 2초 단위로 자른 경우를 seg2, 3초 단위로 자른 경우를 seg3로 표기하였다. 이는 각각 입력 시계열의 길이를 360과 540으로 설정한 것을 의미한다. 두 번째 변수는 합성곱 계층에서의 활성화 함수 종류로, 전통적으로 사용되는 ReLU를 적용한 조건과, 실험적으로 softmax를 합성곱 계층 내부에 적용한 조건을 각각 relu, softmax로 구분하였다. 이렇게 구성된 조합은 seg2, relu, seg3, softmax 등과 같이 표기되며, <표 5>의 최종 실험표에서는 각 폴드(fold)의 평가 지표와 함께 명시되어 있다. 따라서 각 조건이 모델 성능에 미친 영향을 비교하면서, 실험 결과가 가지는 의미와 시사점을 함께 분석할 것이다.

두 번째 실험은 첫 번째 실험에서 가장 좋은 성능을 보인 모델을 활용하여 다중 분류의 성능을 확인하고자 하였다. 본 연구에서 MFCC 및 Softmax의 과업 의존성 성능 평가를 목표로 하고자 했기 때문에 (1) MFCC 여부와 (2) CNN 입력단에서의 Softmax 적용 여부를 확인하여 그 성능을 확인하고자 했

다. 따라서 해당 실험에서 MFCC와 Softmax가 가지는 임상적 분석에서의 평가를 추가 분석하고자 한다.

마지막 실험은 두 번째 실험과 동일하게 첫 번째 실험에서 가장 좋은 성능을 보인 모델을 기준으로 이진 분류를 진행할 때, Mel Filter Bank 수의 변화에 따른 성능을 확인함으로써 본 연구에서 MFCC 임상적 적용 가능성과 데이터 차원 변화 시, 데이터 축소의 가능성을 판단하고자 했다. 이를 통해 경량화 가능성과 MFCC 실질적 도입 가능성에 대해 분석할 것이다.

4.1 1차 실험: 이진 분류

4.1.1 MFCC 기반의 convLSTM

MFCC 기반의 convLSTM 모델은 시계열 생체신호 분석에 있어서 주파수 정보를 활용한 입력과 시공간 특성 학습이 결합된 구조로, 본 연구의 핵심 실험 모델 중 하나다. 이 모델은 MFCC를 통해 ECG 신호의 시간-주파수 변환 특성을 전처리하여 입력으로 사용하고, 이후 convLSTM 구조를 통해 국소적 특징 추출과 순차적 시점 간 연관성을 동시에 학습하도록 설계되었다. 본 단락에서는 해당 모델이 네 가지 조건 조합(seg2/seg3, softmax/relu)에 따라 어떤 성능 차이를 보였는지를 분석하고, 이를 통해 MFCC와 활성화 함수 선택의 영향력을 논의하고자 한다.

실험 결과에 따르면, 전체적으로 가장 우수한 성능은 seg3, softmax 조건에서 나타났다. 이 조합은 5-Fold 평균 F1-score 0.8516, 평균 Accuracy 0.9200으로 측정되어, 동일 모델 내 조건들 중 가장 뛰어난 결과를 보였다. 이는 ECG 신호를 3초 단위로 나누었을 때, 리듬 흐름이 충분히 포함되어 모델이

안정적인 패턴을 학습할 수 있었고, softmax 활성화 함수를 통해 필터 간 경쟁 메커니즘이 작동하여 주요 특징이 선별적으로 강화되었기 때문에 해석할 수 있다. 특히 softmax를 conv 레이어에 적용한 본 연구의 시도는, 일반적인 ReLU 사용과는 다른 특이한 접근이었으며, 이러한 실험적 도전이 의미 있는 결과로 이어졌다는 점에서 학문적 기여도를 가진다.

반면, seg2, relu 조건은 평균 F1-score가 0.7062로 다소 낮은 결과를 보였다. 특히 Fold2에서의 F1-score가 0.0000으로 기록된 점은 성능 불안정성을 시사한다. 이는 상대적으로 짧은 2초 구간 내에서 정상과 비정상 리듬의 구분을 학습하기에 정보량이 부족했고, ReLU의 비선형성 특성상 특정 필터가 과도하게 활성화되거나 사라지는 'dead neuron' 현상이 발생했을 가능성을 시사한다. 반대로, 동일한 seg2에서도 softmax 조건에서는 F1-score가 0.8203으로 크게 향상되었는데, 이는 softmax가 다양한 주파수 패턴 중 중요도를 자동으로 조절하고, 필터 간 정보 표현의 균형을 도모하는 데 긍정적인 영향을 주었음을 의미한다.

또한 seg3 조건에서는 ReLU 역시 일정 수준 이상의 성능을 보였으며 (F1-score 0.8301), 이는 신호가 충분히 길 경우에는 ReLU의 간결한 활성화 메커니즘이 모델 복잡도를 억제하면서도 안정적인 학습을 가능케 한다는 해석도 가능하다. 그러나 그럼에도 불구하고 softmax 조건이 평균적으로 더욱 일관되고 높은 성능을 보였다는 점에서, 본 연구에서 제안한 softmax 기반 활성화 전략은 주목할 만하다.

이러한 결과는 MFCC의 주파수 기반 특징 표현이 convLSTM 구조 내에서 효과적으로 활용될 수 있다는 것을 보여주며, 특히 softmax 활성화와 결합할 경우 필터 간 정보를 더욱 정제된 형태로 전달할 수 있음을 시사한다. MFCC와 softmax의 결합이 ECG 신호에서 주요 리듬 패턴을 더욱 선명하게 부각시켜줌으로써, 해당 모델이 고성능 분류기를 구성할 수 있었던 핵심 배경이라고

볼 수 있다.

<표 1. MFCC 기반의 convLSTM 모델 결과표>

	Accuracy	Precision	Recall	F1-score	Specificity
fold = 1					
seg2, relu	0.9065	0.9658	0.8237	0.8891	0.9756
seg2, softmax	0.8416	0.8883	0.7098	0.7891	0.936
seg3, relu	0.9038	0.9222	0.8404	0.8794	0.9492
seg3, softmax	0.8687	0.8982	0.8026	0.8477	0.924
fold = 2					
seg2, relu	0.5446	0	0	0	1
seg2, softmax	0.8631	0.8181	0.8641	0.8405	0.8623
seg3, relu	0.7383	0.8242	0.4741	0.602	0.9275
seg3, softmax	0.8754	0.8918	0.8266	0.8579	0.9162
fold = 3					
seg2, relu	0.9311	0.9245	0.9242	0.9244	0.937
seg2, softmax	0.8546	0.8375	0.8084	0.8227	0.8877
seg3, relu	0.9191	0.9418	0.8593	0.8986	0.9619
seg3, softmax	0.8658	0.8782	0.8188	0.8475	0.9051
fold = 4					
seg2, relu	0.9311	0.9479	0.8981	0.9223	0.9587
seg2, softmax	0.8516	0.8274	0.8144	0.8208	0.8783
seg3, relu	0.9052	0.9148	0.8522	0.8824	0.9431
seg3, softmax	0.8764	0.8875	0.8344	0.8601	0.9116
fold = 5					
seg2, relu	0.5809	0.5748	0.3056	0.399	0.811
seg2, softmax	0.864	0.874	0.7878	0.8287	0.9186
seg3, relu	0.9113	0.9381	0.8432	0.8881	0.9601
seg3, softmax	0.8649	0.8854	0.8079	0.8449	0.926

4.1.2 MFCC 기반의 CNN + BiLSTM hybrid model

본 절에서는 MFCC 기반의 CNN + BiLSTM 하이브리드 모델의 실험 결과를 분석한다. 해당 모델은 CNN을 통한 주파수 기반 특징 추출과 BiLSTM을 통한 시계열 맥락 정보를 함께 처리할 수 있도록 설계되어, ECG 신호처럼 복잡한 시계열 데이터를 분석하기에 매우 적합하다. 특히 MFCC로 전처리된 데이터를 CNN에 먼저 입력함으로써, ECG의 시간-주파수 패턴을 보다 구조화된 형태로 학습시킬 수 있으며, 이후 BiLSTM 층은 앞뒤 맥락을 모두 인식하여 세밀한 리듬의 변화를 포착할 수 있게 된다.

실험은 앞선 모델과 동일하게 총 4가지 조건(seg2, seg3 × relu, softmax 조합)으로 구성되었으며, 그 중 가장 뛰어난 결과는 seg2, softmax 조건에서 확인되었다. 해당 조건은 평균 F1-score 0.93078, 평균 Accuracy 0.94402로, 본 연구의 모든 실험 조건 중에서도 최상위 성능을 기록했다. 이는 3초 구간 내에서 ECG 신호의 리듬 흐름이 충분히 포착되었고, CNN 층에서 softmax 활성화 함수를 통해 채널 간 경쟁이 유도되며 중요한 주파수 패턴이 강조된 것이 주된 원인으로 분석된다.

반면, seg2, relu 조건에서는 평균 F1-score가 0.8785로 다소 낮은 결과를 보였으며, 이는 2초 길이의 신호 구간이 모델이 리듬 패턴을 안정적으로 학습하기에 충분한 정보량을 제공하지 못한 영향일 수 있다. 특히 relu의 경우 중요한 패턴을 선형적으로 통과시키는 구조이기 때문에, 주파수 패턴의 구분이 약한 짧은 세그먼트에서는 성능 저하가 발생할 수 있다. 그러나 이 조건조차도 전체적으로 F1-score 편차가 크지 않았고, 안정적인 결과를 보인다는 점은 MFCC 기반 CNN + BiLSTM 구조의 강력한 일반화 성능을 입증한다.

흥미로운 점은, seg2 조건에서도 softmax가 적용된 경우(F1-score 0.8905, Accuracy 0.9511) 성능이 seg3 조건과 큰 차이를 보이지 않았다는 점이다. 이

는 MFCC 자체가 시간-주파수 정보를 고도로 요약하기 때문에, 짧은 시간 단위의 신호에서도 효과적으로 특징을 추출할 수 있음을 의미하며, softmax가 이 특징을 극대화하는 데 큰 역할을 했음을 뒷받침한다.

또한 relu와 softmax의 차이를 살펴보면, relu가 적용된 경우에도 전반적인 정확도는 높게 유지되었으나, F1-score에서는 미세한 차이를 보였다. 이는 relu가 신호의 크기를 강조하여 학습의 폭을 넓히는 데는 유리하지만, 민감도 (Recall)와 정밀도(Precision) 사이의 균형을 잡기에는 softmax보다 부족할 수 있다는 점을 시사한다.

요약하자면, MFCC + CNN + BiLSTM 모델은 본 연구의 모든 실험 중에서 가장 일관되고 우수한 성능을 보였으며, 특히 softmax 활성화 함수와의 결합 효과가 두드러졌다. 이러한 결과는 CNN이 추출한 주파수 특징을 BiLSTM이 맥락적으로 재구성함으로써 ECG의 불규칙적인 패턴을 효과적으로 파악했음을 나타낸다. MFCC를 통한 주파수 변환 전처리와 softmax 기반 활성화 구조는 신호의 본질적인 구조를 강화하고, 모델이 의미 있는 특징을 선택적으로 학습하는 데 긍정적인 영향을 주었다고 해석할 수 있다.

결과적으로 본 모델은 MFCC 기반 처리, CNN 필터링, BiLSTM의 순차성 인식, 그리고 softmax의 필터 선택적 강조 효과가 복합적으로 작용하며 최고의 성능을 달성했다. 이는 본 논문에서 제시한 softmax 기반 CNN 활용이 기존 방식과 차별화되는 실험적 성과를 보였다는 점에서 중요한 오리지널리티로 평가될 수 있다.

요약하면, MFCC와 softmax 조합은 ECG의 주파수적 특성과 시계열 특성을 동시에 극대화할 수 있는 설계 전략이며, 본 모델은 이러한 전략이 실제 성능 향상에 기여함을 실험적으로 입증해주고 있다. 이는 향후 MFCC 기반 딥러닝 모델 설계 시, convLSTM 구조뿐 아니라 다양한 네트워크 아키텍처에도 적용될 수 있는 가능성을 시사한다.

<표 2. MFCC 기반의 CNN+BiLSTM 모델 결과표>

	Accuracy	Precision	Recall	F1-score	Specificity
fold = 1					
seg2, relu	0.908	0.9361	0.8367	0.8836	0.9591
seg2, softmax	0.942	0.9596	0.8988	0.9282	0.9729
seg3, relu	0.8917	0.9224	0.8321	0.8749	0.9415
seg3, softmax	0.9416	0.9572	0.9126	0.9343	0.9659
fold = 2					
seg2, relu	0.9123	0.9461	0.8375	0.8885	0.9658
seg2, softmax	0.9472	0.9585	0.913	0.9352	0.9717
seg3, relu	0.9115	0.9253	0.8764	0.9002	0.9409
seg3, softmax	0.94	0.9526	0.9137	0.9327	0.962
fold = 3					
seg2, relu	0.9056	0.9104	0.8584	0.8837	0.9395
seg2, softmax	0.9434	0.9508	0.9116	0.9308	0.9662
seg3, relu	0.8479	0.7952	0.897	0.843	0.807
seg3, softmax	0.9416	0.9598	0.9098	0.9341	0.9682
fold = 4					
seg2, relu	0.8952	0.9464	0.7937	0.8634	0.9678
seg2, softmax	0.9447	0.9691	0.896	0.9311	0.9796
seg3, relu	0.9184	0.9403	0.8764	0.9072	0.9535
seg3, softmax	0.9272	0.9452	0.8919	0.9178	0.9568
fold = 5					
seg2, relu	0.8967	0.8922	0.8599	0.8737	0.9259
seg2, softmax	0.9428	0.9687	0.8918	0.9286	0.9793
seg3, relu	0.9081	0.9243	0.8694	0.896	0.9405
seg3, softmax	0.9375	0.9415	0.9199	0.9306	0.9522

4.1.3 convLSTM

MFCC 전처리 없이 원본 ECG 신호를 직접 입력하여 학습한 convLSTM 모델의 실험 결과를 분석한다. 해당 모델은 MFCC 없이도 시계열 패턴을 학습할 수 있는지 여부와, softmax 활성화 함수의 유의미한 효과가 MFCC 유무와 관계없이 나타나는지를 확인하기 위한 비교 실험의 일환으로 설계되었다. 모델 구조 자체는 4.1.1의 MFCC + convLSTM과 동일하며, 입력의 차이만 존재한다.

먼저, 전반적인 성능을 살펴보면, MFCC가 적용되지 않았음에도 불구하고 convLSTM 모델은 상당히 견고한 결과를 보였다. 특히 seg3, softmax 조건에서는 평균 Accuracy가 0.9537, 평균 F1-score가 0.9024로 나타나, MFCC가 있는 모델과 비교해도 크게 뒤처지지 않는 성능을 기록했다. 이는 원본 신호의 길이가 충분할 경우, convLSTM 구조 자체의 시공간 특징 추출 능력이 효과적으로 작동했음을 의미한다. 긴 세그먼트는 시간축상에서의 리듬성과 전이 구간을 충분히 포착할 수 있어, MFCC 없이도 의미 있는 학습이 가능하다는 점을 보여준다.

다만 MFCC가 있었던 경우에 비해 전체적으로 F1-score 편차가 다소 존재하며, softmax 조건이 아닌 경우엔 성능이 보다 불안정해지는 경향이 뚜렷했다. 예를 들어 seg2, relu 조건에서는 평균 F1-score가 0.8866으로, seg3 대비 약간 낮은 수치를 보였다. 이는 2초 길이의 세그먼트가 ECG의 리듬 패턴 전체를 담기에는 부족했고, 주파수 기반 요약 정보 없이 단순 파형 입력만으로는 모델이 핵심 특징을 구별하는 데 어려움을 겪었기 때문으로 해석된다.

특히 흥미로운 점은 softmax가 적용된 조건들에서 성능이 상대적으로 우수했다는 점이다. seg2, softmax 조건에서도 평균 F1-score는 0.9000 이상으로, MFCC 없이도 softmax의 필터 선택 메커니즘이 주요 ECG 특징을 강조하는

데 도움을 주었음을 시사한다. 이는 softmax가 conv 레이어에서 각 필터의 출력을 정규화하여 상대적으로 더 중요한 특징을 강조하는 구조로 작용하고, 이러한 필터 간 경쟁이 신호의 주요 부분을 자동으로 선택하게 하는 데 유효하게 작용했기 때문으로 볼 수 있다.

결과적으로, MFCC가 없는 convLSTM 모델은 MFCC가 있는 경우에 비해 전반적으로 성능이 조금 낮고 변동성도 컸지만, softmax를 적용할 경우 그 격차를 줄이는 데 일정 부분 성공하였다. 이는 본 연구에서 제시하는 softmax의 가능성을 강조해주는 중요한 실험 결과로, 특히 고정된 feature extractor 없이도 softmax가 일정 수준의 패턴 강조 효과를 가질 수 있음을 시사한다.

따라서 MFCC가 없는 상태에서도 convLSTM은 일정 수준 이상의 성능을 유지할 수 있었고, 특히 softmax의 적용이 그 성능을 견고하게 만들어주는 데 크게 기여했다. 본 결과는 MFCC 없이도 softmax가 모델의 주목 특성을 향상시키는 데 유효하다는 점을 실험적으로 입증하며, softmax의 일반적 가능성과 확장성에 대한 중요한 단서가 된다.

<표 3. ConvLSTM 모델 결과표>

	Accuracy	Precision	Recall	F1-score	Specificity
fold = 1					
seg2, relu	0.9437	0.8758	0.758	0.8127	0.9793
seg2, softmax	0.8693	0.7443	0.2884	0.4157	0.981
seg3, relu	0.9531	0.9291	0.7661	0.8398	0.9888
seg3, softmax	0.8524	0.6614	0.1627	0.2612	0.9841
fold = 2					
seg2, relu	0.9442	0.8854	0.7511	0.8127	0.9813
seg2, softmax	0.8732	0.7986	0.2855	0.4206	0.9862
seg3, relu	0.8895	0.8404	0.3836	0.5267	0.9861
seg3, softmax	0.8514	0.6138	0.1965	0.2977	0.9764
fold = 3					
seg2, relu	0.9381	0.9637	0.6405	0.7695	0.9954
seg2, softmax	0.8725	0.7126	0.3515	0.4708	0.9727
seg3, relu	0.9413	0.8787	0.7352	0.8006	0.9806
seg3, softmax	0.8595	0.6215	0.3161	0.419	0.9632
fold = 4					
seg2, relu	0.9468	0.9301	0.7245	0.8145	0.9895
seg2, softmax	0.8697	0.7291	0.3058	0.4309	0.9781
seg3, relu	0.9498	0.8962	0.7766	0.8322	0.9828
seg3, softmax	0.8659	0.7233	0.2653	0.3882	0.9806
fold = 5					
seg2, relu	0.948	0.9453	0.7194	0.817	0.992
seg2, softmax	0.8663	0.6872	0.3146	0.4316	0.9725
seg3, relu	0.8397	0	0	0	1
seg3, softmax	0.8505	0.7118	0.1131	0.1951	0.9913

4.1.4 CNN + BiLSTM hybrid model

본 절에서는 MFCC 전처리를 적용하지 않고, 원본 ECG 신호를 직접 입력값으로 사용하여 구성된 CNN + BiLSTM 모델의 실험 결과를 분석한다. 해당 모델은 CNN을 활용하여 국소적인 신호 특징을 추출하고, 이후 BiLSTM을 통해 양방향 시점의 정보를 통합하여 시간 축상의 흐름을 학습하도록 설계되었다. MFCC 없이도 성능이 확보되는지, 그리고 softmax 활성화가 주는 효과가 MFCC의 부재 하에서도 유의미한지 등을 확인하는 것이 본 절의 핵심이다.

실험 결과에 따르면, CNN + BiLSTM (no MFCC) 모델은 전체적으로 뛰어난 안정성을 보이며 높은 성능을 기록하였다. 특히 seg2, relu 조건에서는 평균 Accuracy가 0.9641, 평균 F1-score가 0.9148로 측정되어, MFCC를 사용한 경우와 비교해도 전혀 손색이 없는 결과를 나타냈다. 이는 CNN의 강력한 로컬 필터링 능력이 원본 신호의 유효한 구간에서 주요 특징을 효과적으로 추출했고, BiLSTM의 시점 정보 통합이 시계열적 맥락을 안정적으로 반영했음을 보여준다. 즉, MFCC 없이도 모델 자체의 구조적 강점이 일정 수준 이상의 성능을 담보할 수 있다는 점에서 의미 있는 결과로 해석된다.

활성화 함수의 측면에서도 흥미로운 경향이 관찰되었다. relu 조건은 전반적으로 높은 정확도와 F1-score를 보였으나, softmax를 적용한 경우에도 큰 성능 저하는 나타나지 않았다. 예를 들어 seg3, softmax 조건은 평균 F1-score 0.9055, Accuracy 0.9516으로, relu 조건과 유사하거나 오히려 더 고른 결과를 보여주었다. 이는 softmax가 CNN 필터 간 경쟁을 유도하여 특정 구간에 집중적인 특성 추출을 유도하는 방식이, MFCC 없이도 일정 수준의 패턴 강조 효과를 갖는다는 점을 보여준다.

반면 seg2, softmax 조건에서는 상대적으로 다소 낮은 평균 F1-score (0.8999)를 기록했으며, 이는 짧은 세그먼트 구간 내에서는 softmax의 특성상

일부 중요 특징만 선택적으로 강조되며, 전체 패턴을 포괄하지 못한 가능성을 시사한다. 그러나 이 조건 역시 Accuracy 0.9515로 높은 정확도를 유지한 것으로 보아, 실제 분류의 예측력은 유지되었고, 다만 정밀도와 재현율 간의 균형(F1-score) 측면에서 relu보다 불리했음을 알 수 있다.

전반적으로 CNN + BiLSTM 구조는 MFCC 없이도 강한 일반화 능력을 보였으며, raw ECG 신호로도 뛰어난 성능을 구현할 수 있음을 입증했다. 특히 softmax의 효과가 여전히 유의미하게 나타났다는 점은 본 연구에서 제시하는 “softmax 기반 CNN 특성 강화” 접근이 단순히 MFCC에 한정된 것이 아니라, 다양한 입력 조건에서도 적용가능한 전략임을 시사한다. 이는 향후 생체신호나 시계열 신호 처리에 있어, 보다 유연한 적용 가능성을 뒷받침하는 결과다.

<표 4. CNN+BiLSTM 모델 결과표>

	Accuracy	Precision	Recall	F1-score	Specificity
fold = 1					
seg2, relu	0.9634	0.9165	0.8505	0.8823	0.9851
seg2, softmax	0.962	0.9489	0.8077	0.8727	0.9916
seg3, relu	0.9675	0.9288	0.8632	0.8948	0.9874
seg3, softmax	0.9639	0.9361	0.8318	0.8808	0.9892
fold = 2					
seg2, relu	0.9653	0.9578	0.8213	0.8843	0.993
seg2, softmax	0.9645	0.9517	0.8216	0.8819	0.992
seg3, relu	0.9634	0.9176	0.8477	0.8812	0.9855
seg3, softmax	0.9632	0.9251	0.8383	0.8796	0.987
fold = 3					
seg2, relu	0.9657	0.9414	0.8396	0.8876	0.99
seg2, softmax	0.9618	0.9443	0.8111	0.8726	0.9908
seg3, relu	0.966	0.9109	0.8737	0.8919	0.9837
seg3, softmax	0.9655	0.8999	0.8831	0.8914	0.9813
fold = 4					
seg2, relu	0.9644	0.9358	0.8367	0.8835	0.989
seg2, softmax	0.9619	0.9428	0.8129	0.8731	0.9905
seg3, relu	0.9623	0.9566	0.8014	0.8721	0.993
seg3, softmax	0.9627	0.9444	0.8152	0.8751	0.9908
fold = 5					
seg2, relu	0.9645	0.9296	0.844	0.8847	0.9877
seg2, softmax	0.9587	0.9115	0.8239	0.8655	0.9846
seg3, relu	0.9695	0.9509	0.8538	0.8997	0.9916
seg3, softmax	0.9666	0.9304	0.8555	0.8914	0.9878

<표 5. 최종 결과 종합표>

	AVG_F1	AVG_AUC
MFCC + convLSTM		
seg2, relu	0.62696	0.77884
seg2, softmax	0.82036	0.85498
seg3, relu	0.8301	0.87554
seg3, softmax	0.85162	0.87024
MFCC + CNN_BiLSTM		
seg2, relu	0.87858	0.90356
seg2, softmax	0.93078	0.94402
seg3, relu	0.88426	0.89552
seg3, softmax	0.9299	0.93758
convLSTM		
seg2, relu	0.80528	0.94416
seg2, softmax	0.43392	0.8702
seg3, relu	0.59986	0.91468
seg3, softmax	0.31224	0.85594
CNN_BiLSTM		
seg2, relu	0.88448	0.96466
seg2, softmax	0.87316	0.96178
seg3, relu	0.88794	0.96574
seg3, softmax	0.88366	0.96438

4.2. 2차 실험: 다중 분류

4.2.1 실험 결과

MFCC 필터뱅크 수를 40, 64, 128, 256으로 조정하여 수행한 이진 분류 실험 결과는 다음과 같다. 주요 성능 지표인 F1-score를 기준으로 비교하였으며, 부가적으로 Accuracy를 참고하였다. 실험 결과, 필터뱅크 수가 40일 때 가장 높은 F1-score를 기록하였으며, 필터뱅크 수가 과도하게 증가하는 경우(256)에는 미세한 성능 저하가 관찰되었다.

아래의 표는 필터뱅크 수별 성능 지표를 요약한 것이다.

모든 파라미터에서 유사한 수준의 높은 정확도가 확인되었지만, n_mels 수가 증가함에 따라 오히려 성능이 미세하게 감소하는 경향이 반복적으로 관찰되었다. 이는 MFCC가 mel 스펙트로그램의 전체 정보를 입력받는 것이 아니라, 주성분 추출을 통해 일부 정보만을 사용하는 구조적 특성 때문으로 해석할 수 있다. 즉, filter bank 수가 지나치게 많아질 경우, 실제 모델이 활용하지 않는 고주파 정보까지 포함되면서 오히려 특징의 명확성이 저하되거나, 불필요한 변동성이 유입되어 일반화 성능에 부정적 영향을 줄 수 있다.

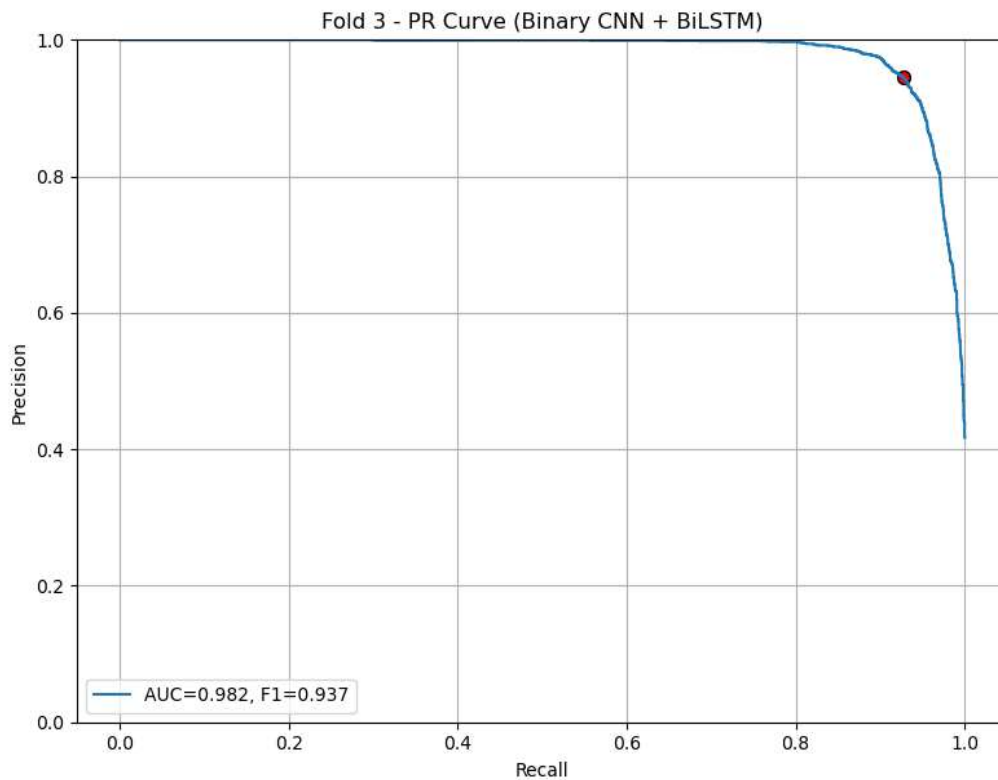
또한, n_mels 수를 줄였을 때에도 MFCC 20개만으로 성능 저하 없이 효과적인 분류 결과를 도출할 수 있다는 점은 특징 압축의 효율성을 뒷받침하며, 모델 경량화 관점에서도 중요한 시사점을 제공한다. 실제로 n_mels=40 설정의 모델은 n_mels=256 설정 대비 약 0.006의 F1-score 향상을 보이며, 연산 복잡도 측면에서도 더 유리한 결과를 나타냈다.

따라서 본 실험을 통해 MFCC 기반 이진 분류 모델에서의 filter bank 수 설정은 성능에 직접적인 영향을 미칠 수 있으며, 무작정 높은 해상도의 필터

를 사용하는 것이 항상 최선은 아님을 확인할 수 있었다. 특히, MFCC 수가 고정된 상태에서는 적정 수준의 n_mels 설정이 오히려 정보 손실 없이 모델의 해석 가능성과 효율성을 모두 확보하는데 유리한 선택이 될 수 있음을 시사한다.

<표 6. Filter Bank = 40 결과>

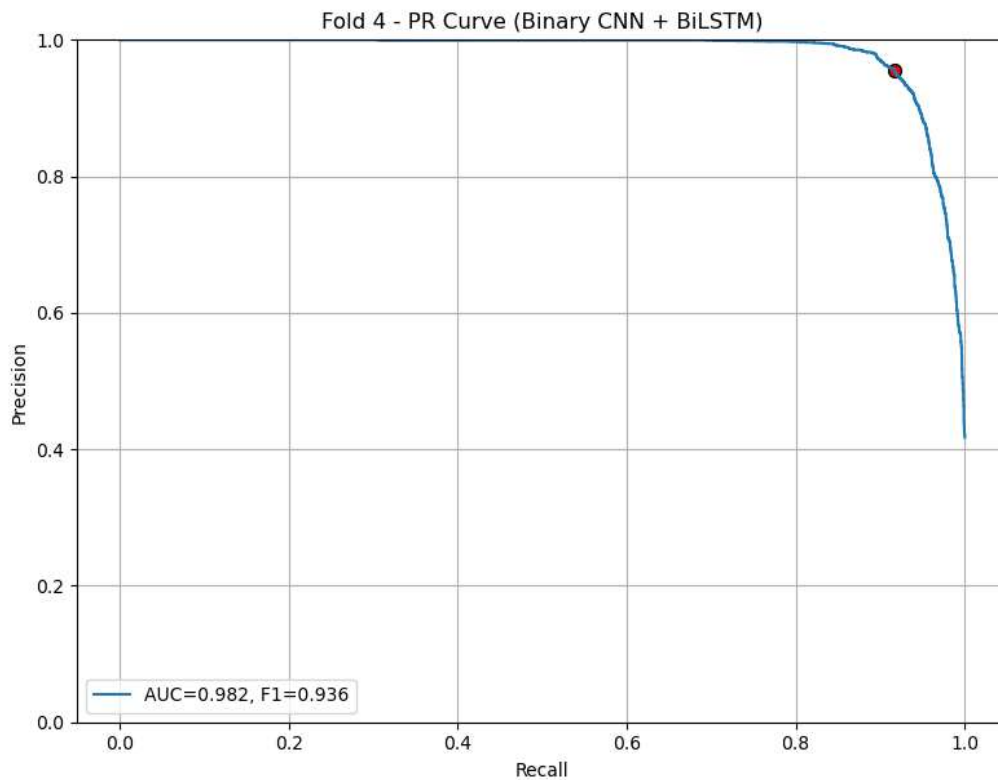
	precision	recall	f1-score	support
Class 0	0.9479	0.9621	0.955	4940
Class 1	0.946	0.9263	0.936	3539
Accuracy			0.9472	8479
macro-avg	0.947	0.9442	0.9455	8479
weighted-avg	0.9471	0.9472	0.9471	8479



<그림 7> Filter Bank = 40 PR-Curve>

<표 7. Filter Bank = 64 결과>

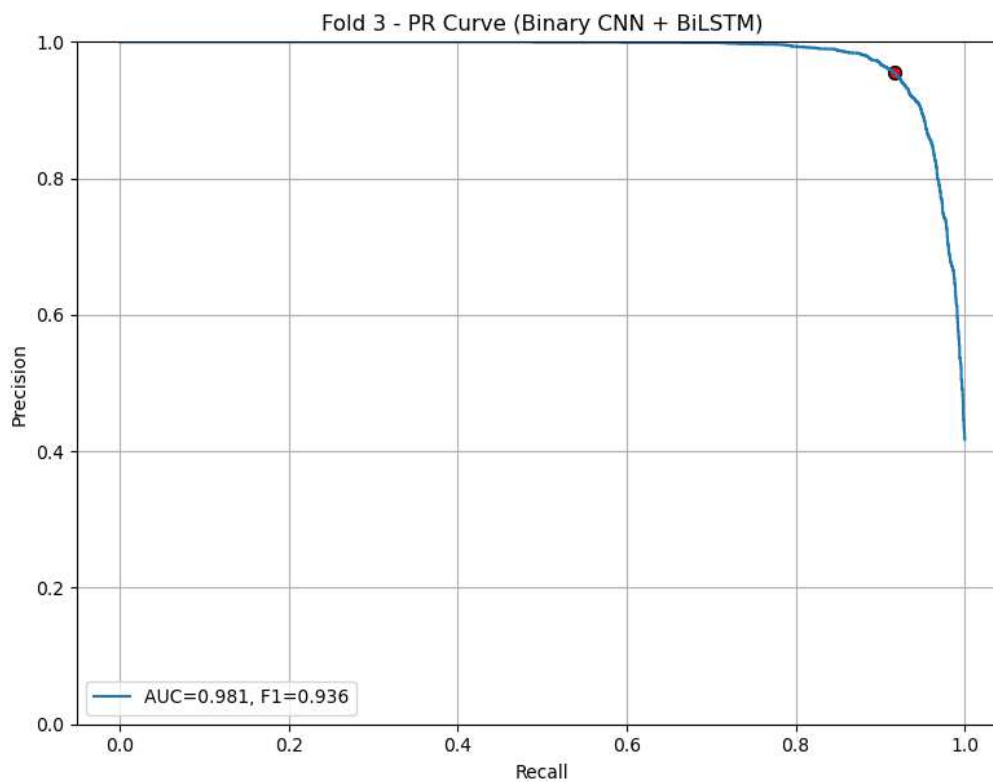
	precision	recall	f1-score	support
Class 0	0.9335	0.9779	0.9552	4940
Class 1	0.967	0.9028	0.9338	3539
Accuracy			0.9466	8479
macro-avg	0.9503	0.9404	0.9445	8479
weighted-avg	0.9475	0.9466	0.9463	8479



<그림 8> Filter Bank = 64 PR-Curve>

<표 8. Filter Bank = 128 결과>

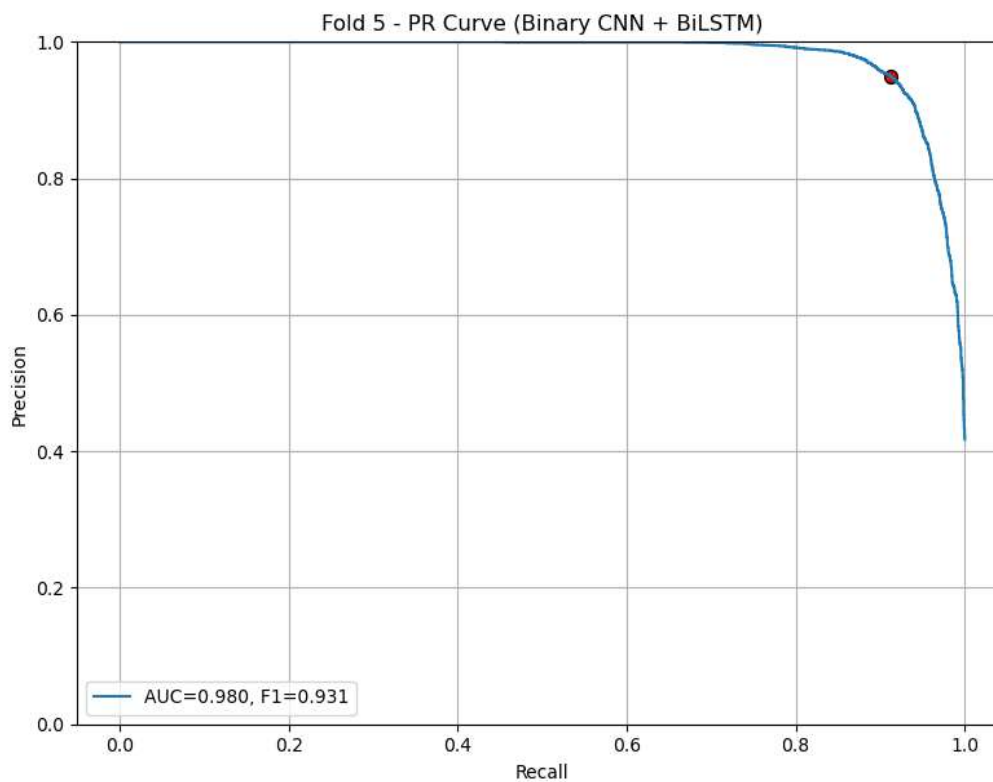
	precision	recall	f1-score	support
Class 0	0.943	0.9672	0.9549	4940
Class 1	0.9525	0.9183	0.9351	3539
Accuracy			0.9468	8479
macro-avg	0.9477	0.9428	0.945	8479
weighted-avg	0.947	0.9468	0.9467	8479



<그림 9> Filter Bank = 128 PR-Curve

<표 9. Filter Bank = 256 결과>

	precision	recall	f1-score	support
Class 0	0.9455	0.9551	0.9502	4940
Class 1	0.9364	0.9231	0.9297	3539
Accuracy			0.9417	8479
macro-avg	0.9409	0.9391	0.94	8479
weighted-avg	0.9417	0.9417	0.9417	8479



<그림 10> Filter Bank = 256 PR-Curve

4.3. 3차 실험: Mel Filter Bank 수 변화

4.3.1 실험 결과

본 실험에서는 다중 클래스 분류에서 MFCC의 효용성을 평가하기 위해, CNN+BiLSTM 모델 구조를 고정한 상태에서 MFCC 적용 여부 및 합성곱 계층의 활성화 함수(ReLU, Softmax) 조합에 따른 네 가지 실험을 설계했다. 모든 실험은 5-fold 교차검증을 통해 수행했으며, 평가지표는 macro-F1 score, weighted-F1 score, accuracy로 구성했다. 또한 정밀도-재현율(PR) 곡선을 기반으로 클래스별 분류 특성을 시각화하여 비교 분석을 병행하고자 했다. 아래의 결과표와 그래프는 각 실험 조건에서 가장 높은 성능을 보인 fold의 결과를 기준으로 제시했다.

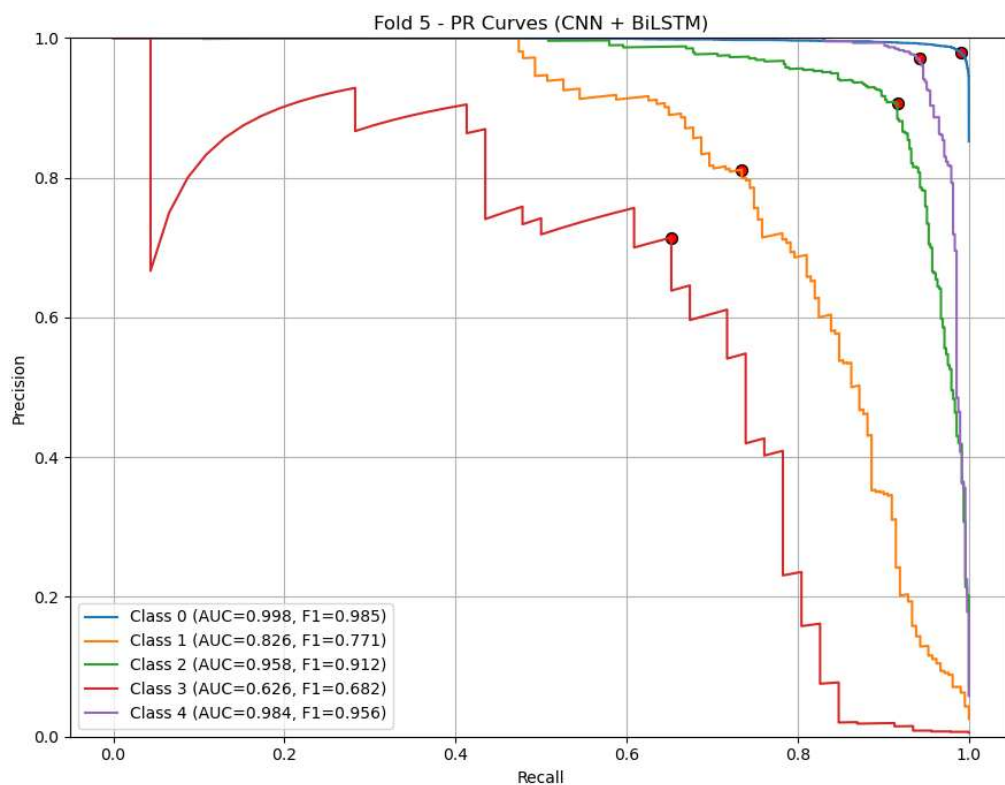
정량적 비교 결과, 네 가지 조건 중 MFCC를 사용하지 않은 CNN+BiLSTM 하이브리드 모델이 가장 우수한 성능을 기록하였고, 뒤이어 MFCC를 적용한 CNN+BiLSTM 모델이 근소한 차이로 두 번째 성능을 나타냈다. 특히 두 모델 간 평균 weighted-F1 score 차이는 약 0.004 수준으로, 이는 환경 변화나 데이터 특성에 따라 충분히 변동 가능한 범위로 해석되며, 단순히 MFCC를 적용한 모델의 성능이 낮다고 일반화하기에는 무리가 있다고 판단했다.

보다 심화된 비교를 위해 PR 곡선을 분석한 결과, MFCC를 적용한 모델이 AAMI 기준 Class 4(Q)에 대해 일관적으로 높은 정밀도와 재현율을 보이며, 해당 클래스의 탐지 성능이 강화됨을 확인할 수 있었다. 구체적으로 해당 클래스에 대해서는 0.99 이상의 Accuracy 및 0.97 이상의 F1-score를 기록하여, MFCC가 특정 이상 신호 패턴을 보다 효과적으로 감지하는 데 기여했음을 시사한다.

다만, AAMI 수퍼 클래스 기준 Q는 원래 진단적 기준이 명확하지 않거나 잡음이 포함된 신호들이 포함되는 범주로, 실제 임상적 판단에서의 해석은 제한적일 수 있다. 이러한 점을 고려하면, 본 모델을 통해 해당 클래스를 추출한 나머지 네 개의 주요 부정맥 클래스(N, S, V, F)에 대한 추가 실험을 통해 보다 임상적으로 유의미한 결과를 확보할 가능성이 크다고 판단된다 [15-16]. 이는 향후 MFCC의 적용 여부를 결정할 때, 단순한 성능 수치 외에도 탐지하고자 하는 대상 클래스의 임상적 중요도를 함께 고려할 필요가 있음을 보였다.

<표 10. CNN+BiLSTM 결과>

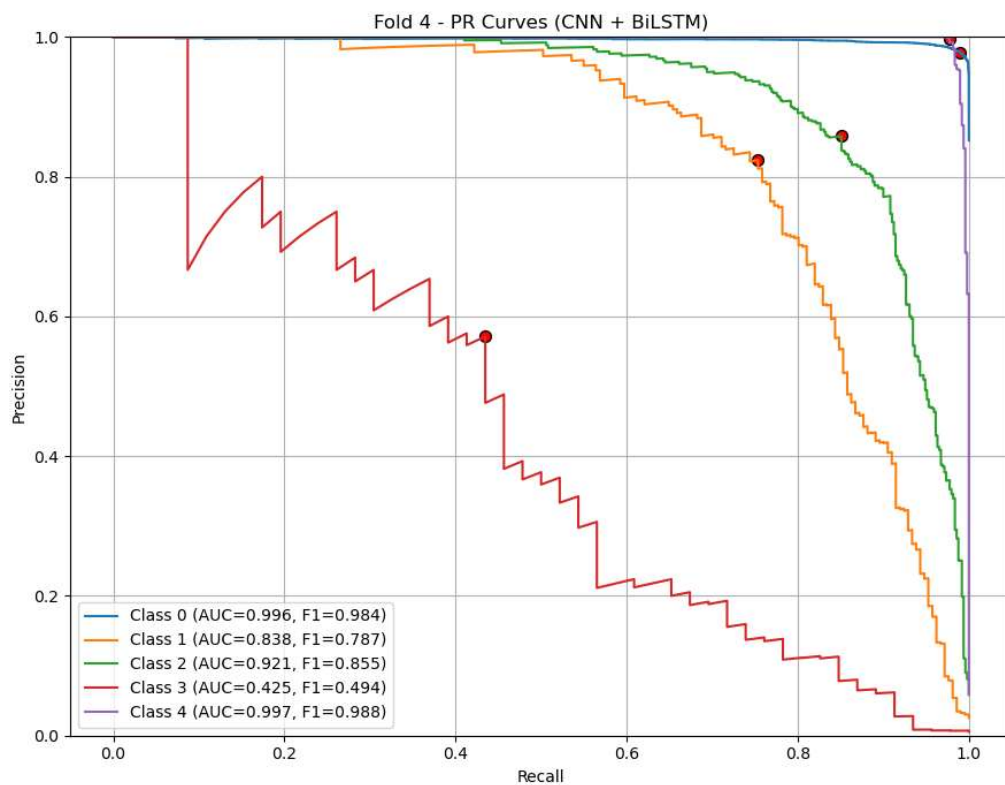
	precision	recall	f1-score	support
Class 0	0.9802	0.9900	0.9851	7105
Class 1	0.7861	0.7488	0.7670	211
Class 2	0.9244	0.8980	0.9100	490
Class 3	0.8800	0.4783	0.6197	46
Class 4	0.9804	0.9261	0.9525	487
Accuracy			0.9719	8339
macro-avg	0.9098	0.8082	0.8469	8339
weighted-avg	0.9714	0.9719	0.9712	8339



<그림 11> CNN+BiLSTM PR-Curve

<표 11. MFCC+CNN+BiLSTM 결과>

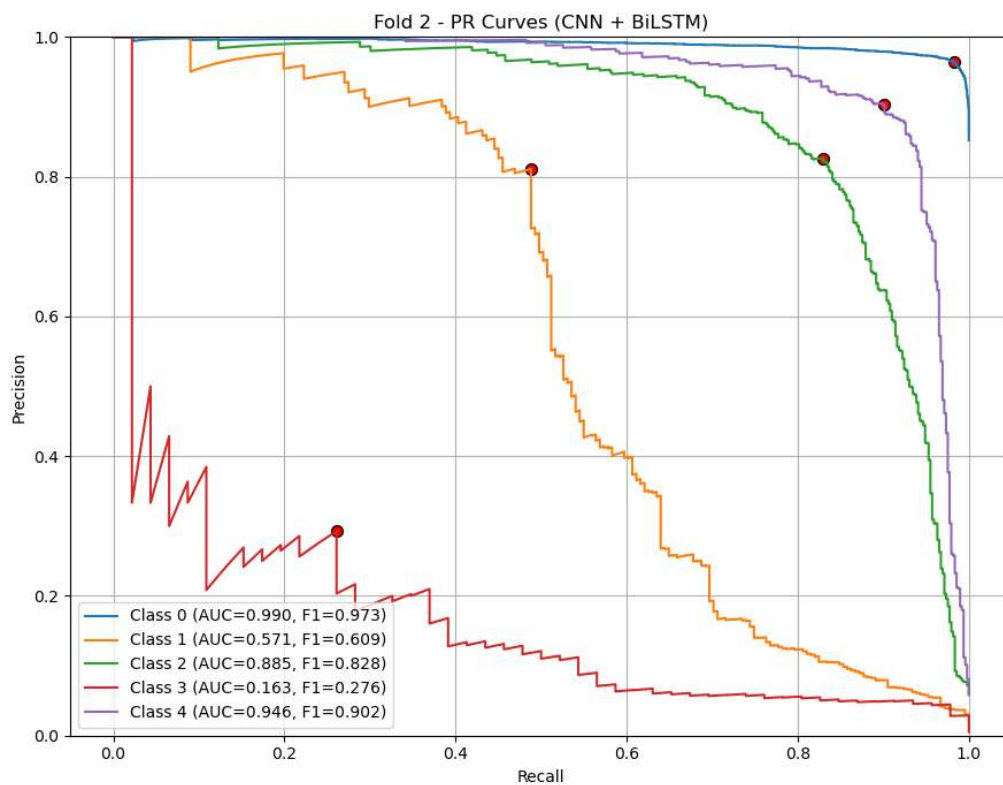
	precision	recall	f1-score	support
Class 0	0.9787	0.9882	0.9834	7105
Class 1	0.8353	0.6730	0.7454	211
Class 2	0.8454	0.8592	0.8522	490
Class 3	0.6250	0.2174	0.3226	46
Class 4	0.9792	0.9671	0.9731	487
Accuracy			0.9671	8339
macro-avg	0.8527	0.7410	0.7754	8339
weighted-avg	0.9653	0.9671	0.9654	8339



<그림 12> MFCC+CNN+BiLSTM PR-Curve

<표 12. CNN+BiLSTM+softmax 결과>

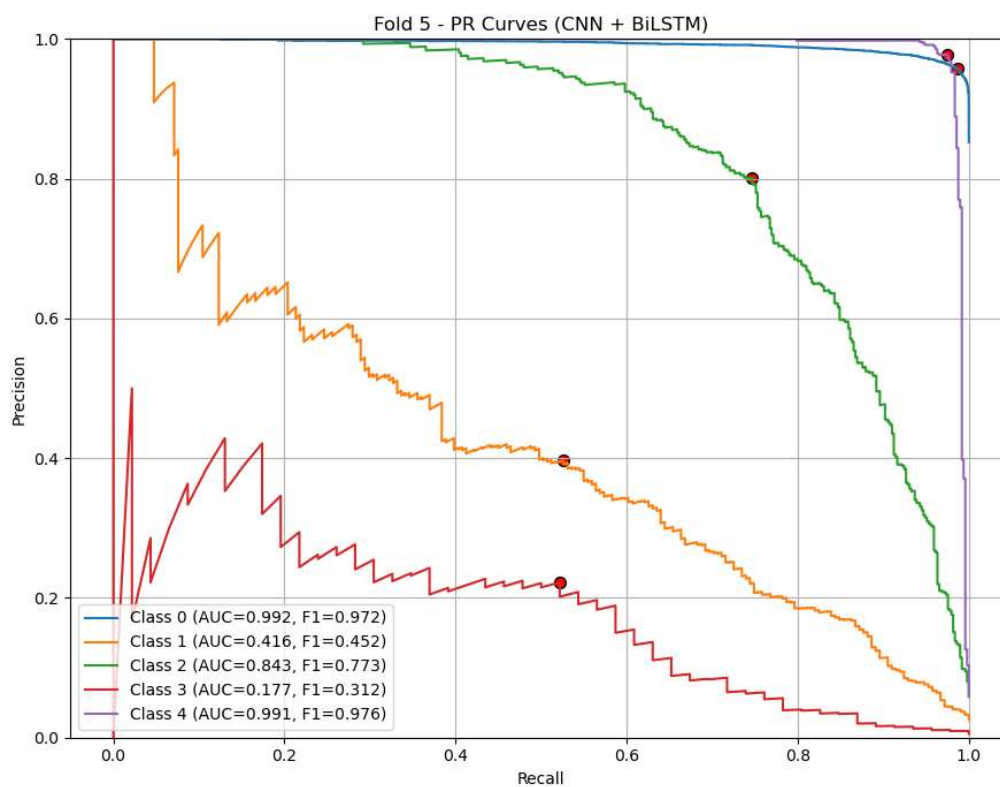
	precision	recall	f1-score	support
Class 0	0.9577	0.9873	0.9723	7105
Class 1	0.8482	0.4502	0.5882	211
Class 2	0.8753	0.7751	0.8221	490
Class 3	0	0	0	46
Class 4	0.9167	0.8809	0.8984	487
Accuracy			0.9496	8339
macro-avg	0.7196	0.6187	0.6562	8339
weighted-avg	0.9424	0.9496	0.9441	8339



<그림 13> CNN+BiLSTM+softmax PR-Curve

<표 13. MFCC+CNN+BiLSTM+softmax>

	precision	recall	f1-score	support
Class 0	0.9520	0.9900	0.9706	7105
Class 1	0.6250	0.1659	0.2622	211
Class 2	0.8266	0.7102	0.7640	490
Class 3	0	0	0	46
Class 4	0.9873	0.9589	0.9729	487
Accuracy			0.9454	8339
macro-avg	0.6782	0.5650	0.5939	8339
weighted-avg	0.9331	0.9454	0.9353	8339



<그림 14> MFCC+CNN+BiLSTM+softmax PR-Curve

5. 고찰

5.1. 1차 실험: 이진 분류

본 절에서는 모델 간 평균 성능 비교 대신, 실험에 사용된 주요 조건 - 세그먼트 길이(seg2/seg3), 활성화 함수 유형(relu/softmax) -에 따른 전반적인 성능 경향을 분석한다. 이는 단순한 평균값 비교에서 드러나지 않는 각 조건 조합의 미세한 영향을 더욱 정밀하게 해석하고, 특히 본 논문이 제안하는 MFCC 기반 하이브리드 구조와 softmax의 결합 효과가 어떤 방식으로 분류 성능에 영향을 주었는지를 보다 명확히 조망하기 위한 것이다.

5.1.1 세그먼트 길이의 영향

실험에 사용된 ECG 데이터는 2초(seg2) 또는 3초(seg3) 단위로 분할되었으며, 이 구간 길이는 모델이 수용할 수 있는 리듬 패턴의 길이와 복잡도에 직접적인 영향을 미쳤다 [17]. 세그먼트를 나누어 연산한 것은, MFCC를 사용한 경우 Seg2가 좋았고, 사용하지 않은 경우는 seg3가 더 좋았다.

이는 다양한 시사점이 있다. MFCC를 활용한 정보를 처리하게 되면, 기존의 신호보다 특징점을 가져가는 방식으로 인해, 시계열적 정보 특성이 부족한 경우가 있어, 긴 범위에 대한 정보로 특징점을 인식하는 것보다 비교적 짧은 주기에서의 특징점 분석을 통해 더 많은 정보를 연산하는 것으로 이해할 수 있다는 점이다.

반면, MFCC를 활용하지 않은 경우에서 Seg3가 더 좋은 이유는 CNN이 다

양한 필터를 통해 국소적 특징을 추출할 때, Seg3 구간은 더 긴 주기와 넓은 변동 폭을 포함하기 때문에 더 풍부한 정보로 모델을 학습할 수 있는 것으로 판단했다.

이를 통해 본 논문이 제시하고자 하는 MFCC 도입을 위한 방법은 특징점 분석을 통한 딥러닝 모델 연산에서 비교적 짧은 경우로 나누어 더 많은 샘플로 연산을 하는 것이 길게 나누어 연산하는 것보다 좋은 것으로 확인할 수 있었다.

5.1.2 활성화 함수의 영향

본 연구에서 가장 주목할 만한 결과는, MFCC 처리를 하여 연산했을 때, CNN 계열 구조에서 softmax 활성화 함수를 사용했을 때 대부분의 조건에서 F1-score가 향상되었다는 점이다. 이는 전통적으로 relu가 사용되는 CNN 구조에서의 통념을 전복하는 결과로, softmax의 특이성과 효과를 뚜렷이 보여주는 핵심 근거로 작용한다.

특히 MFCC + CNN + BiLSTM 모델의 seg2 조건에서 softmax를 적용했을 때 F1-score가 0.93078로 실험 전체 중 가장 높은 수치를 기록하였다. 이는 softmax가 CNN 필터 간 경쟁을 유도하여, MFCC로 추출된 주파수-시간 패턴 중 가장 중요한 부분을 부각시키는 데 큰 역할을 했음을 의미한다. 반면 동일 모델의 relu 조건에서는 F1-score가 0.8891로 다소 낮았으며, 이는 relu가 입력의 양적 증폭은 잘 수행하나, 필터 간 차별화를 유도하는 데에는 상대적으로 부족할 수 있음을 시사한다.

또한 MFCC + convLSTM 모델에서도 softmax 조건은 relu 조건 대비 뚜렷한 성능 향상을 보였다. 예를 들어 seg2 조건에서 relu의 F1-score는 0.7062였으나, softmax 조건에서는 0.8203으로 큰 폭의 상승을 보였다. 이러한 결과

는 softmax가 단순히 CNN 구조에만 특화된 활성화 함수가 아니라, 주요 특징을 강조하고 비주요 특징을 억제하는 데 있어 더 보편적인 장점이 있을 수 있음을 시사한다.

흥미롭게도 MFCC를 적용하지 않은 CNN + BiLSTM 모델에서도 softmax 조건은 경쟁력 있는 성능을 보였으며, 이는 softmax의 효과가 MFCC와 독립적으로도 일정 수준 존재함을 보여준다. 다만, softmax의 효과가 가장 극적으로 발현된 경우는 MFCC와 결합되었을 때였으며, 이 조합이 본 연구의 오리 지널리티를 가장 뚜렷하게 부각시켜주는 실험적 결과라 할 수 있다.

5.1.3. MFCC + softmax 기반 하이브리드 구조의 의의

최종적으로 본 연구에서 가장 주목해야 할 지점은 다음과 같다. MFCC 기반 CNN + BiLSTM 모델에 softmax를 적용했을 때의 성능이, 모든 조건 중 단일 최고 성능을 달성했다는 점이다.

이는 단순한 모델 구조 변경이 아니라, 주파수 정보의 정제된 추출(MFCC)과 선택적 패턴 강조(softmax)라는 두 가지 메커니즘이 상호작용하며, 모델의 특징 인식력을 극대화시킨 결과로 해석된다. ReLU가 단순히 ‘존재하는 특징’을 모두 통과시키는 반면, softmax는 ‘가장 강한 특징’을 선별해내는 역할을 하며, 이 차이가 MFCC와 만나면서 더욱 큰 성능 차이를 만들어낸 것이다.

이러한 결과는 기존의 생체신호 처리 관행에서 볼 수 없었던 접근이며, softmax를 CNN 내 활성화 함수로 사용하는 방식이 ECG 데이터와 같이 시계열적 흐름이 중요한 분야에서도 실질적인 성능 향상으로 이어질 수 있음을 실험적으로 입증했다.

5.2. 2차 실험: 다중 분류

해당 절에서는 1차 실험의 최고 성능 모델을 기반으로 다중 분류에서의 성능을 확인하여 임상적인 효과에 대한 분석을 이어가고자 한다. 결과 자체에서의 평균 성능은 MFCC와 Softmax를 활용하지 않은 경우에서 더 좋은 결과를 보였지만, 단순한 수치적 성능 비교를 확인하는 것이 아닌 클래스별 지표 비교를 통한 유의미한 해석과 이를 통한 후속 연구 발전 가능성에 대해 고찰하고자 한다.

5.2.1. MFCC+Softmax 구조의 임상적 효과

실제 결과에서는 MFCC 와 Softmax를 사용하지 않고, CNN+BiLSTM 하이브리드 구조를 활용한 경우, 클래스별 평균 F1-score가 더 좋은 성능으로 확인되었다. 하지만 수치적인 비교가 아닌 임상적인 분석을 이어가 보면 MFCC와 Softmax의 임상적인 효과를 확인할 수 있다.

MFCC를 사용한 경우에서 MIT-BIH의 AAMI 클래스 분류 기준에서 정상과 Q 클래스에 대해 성능이 압도적으로 좋은 것을 확인할 수 있다. 이는 Softmax를 활용한 경우에서도 0.99 이상의 AUC와 0.97 이상의 F1-score 수치를 확인할 수 있다. 이는 임상적인 측면에서 유의미한 결과가 있다고 판단할 수 있다.

이를 통해 추가 확장 연구에서 MFCC와 Softmax를 활용하여 다중 분류 시, 정상 클래스와 Q 클래스를 우선적으로 분류를 진행한 뒤, 남은 3개의 클래스를 다시 비교분석 및 분류를 진행한다면 유의미한 정확도 증가가 있을 것으로 기대해볼 수 있다.

5.3. 3차 실험: Mel Filter Bank 수 변화

3차 실험 또한 1차 실험에서 가장 성능이 좋았던 조합인 MFCC, Softmax, CNN+BiLSTM 하이브리드 모델로 연구를 진행했다. 앞선 실험들에서는 Mel Filter Bank 수 128을 default 값으로 설정해서 실험했다. 이 숫자는 주파수 정보의 차원을 결정하는 수로 그 수가 적어지면 더 낮은 해상도의 정보를 받아오고, 높은 숫자를 사용할수록 더 높은 해상도의 값을 가져오게 된다.

이를 위해 앞서 실험에서 설정한 4개의 값을 변인으로 두어 어떤 변화가 유효했는지 이 결과가 임상적으로 어떤 의미가 있는지 분석하고자 한다.

5.3.1. Mel Filter Bank 감소에 따른 효과

실험 결과 장에서 제시했던 내용처럼 Mel Filter Bank 수가 감소할수록 더 좋은 결과를 볼 수 있었다. 이는 추후 MFCC를 도입한 경우 적은 128의 해상도 보다 낮은 해상도로 설정하여 진행하는 것이 더 좋은 성능을 보일 수 있는 것으로 확인했다. 이는 MFCC가 주는 정보의 재편이 심전도에서 유의미한 차이를 보이는 것에 충분한 정보를 주고 있는 것으로 확인할 수 있었다. 더 나아가 MFCC 처리 방식이 딥러닝을 위한 차원 변화에 유리한 성질을 가지고 있기 때문에 작은 Mel Filter Bank 수로 변환하여 작업을 진행하면 전처리시, 경량화와 실제 도입 환경에서 정보의 처리에 있어 부담을 줄일 수 있는 장점이 있는 것으로 분석했다.

5.4. 연구 결론과 기여

본 연구에서는 ECG(Electrocardiogram) 신호를 보다 효과적으로 분석하기 위한 두 가지 독창적인 시도를 통해, 기존의 전형적인 신경망 기반 접근법을 한 단계 확장할 수 있는 가능성을 모색하였다. 첫 번째로, 음성 신호 처리 분야에서 널리 활용되어온 멜 주파수 켈스트럼 계수(MFCC)를 ECG 신호 분석에 도입하여, 시간영역과 주파수영역 정보를 혼합적으로 취합하는 하이브리드 모델을 제시하였다. 기존의 CNN 기반 ECG 분류 시스템은 시간영역에서의 파형 특성만을 주로 학습하는 경향이 강했으나, 본 연구에서는 MFCC를 통해 스펙트럼 특성을 강화함으로써 잡음이 끼어 있는 ECG 파형에서도 유의미한 정보를 더욱 안정적으로 포착할 수 있었다. 이러한 시도는 ECG가 지닌 다양한 주파수 대역에서의 미세 변화가 임상적으로 중요한 함의를 가질 수 있다는 가정에서 출발하며, 그 결과 다차원 특징을 학습하는 데 유리한 환경을 조성한다. 실제로 MFCC 기반 하이브리드 모델은 전처리 과정에서부터 ECG 파형의 중요한 대역에 집중할 수 있도록 유도하고, 이후 CNN 계층이 해당 대역의 정보를 분류 모델에 맞춰 세밀하게 보완함으로써 더욱 높은 분류 성능과 잡음 내성을 확보하는 데 성공하였다.

두 번째로, 합성곱 신경망(CNN) 구조의 중간 합성곱 계층에서 Softmax 함수를 적용함으로써 채널 간의 경쟁 기제를 명확하게 도입하는 새로운 방법론을 시도하였다. 이는 보편적으로 CNN 설계 시에는 중간 계층에서 ReLU 계열 활성화 함수를 사용하는 것이 관례라는 점에서, 연구자들의 관심을 끌 만한 참신한 접근이다. 본 실험에서는 이러한 방식이 ECG 신호 특유의 노이즈 및 특징적 패턴을 구별하는 데 긍정적 영향을 미친다는 사실이 확인되었다. 채널 간 출력이 경쟁 관계에 놓이면서, 한 채널이 결정적인 패턴을 강하게 포착할

경우 다른 채널에서 발생할 수 있는 중복적 특징이나 불필요한 잡음이 자연스럽게 억제되는 효과가 나타났기 때문이다. 이는 곧, 중간 계층의 표현력이 제한될 것이라는 일반적인 우려와 달리, ECG와 같은 생체 신호 분석에서는 오히려 핵심 패턴을 더 또렷이 부각시켜 정확도를 높이는 결과로 이어질 수 있음을 의미한다.

이처럼 MFCC 기반 하이브리드 구조와 중간 합성곱 계층의 Softmax 적용이라는 두 가지 새로운 방식의 설계는, 각각 독립적으로도 의미 있지만 상호 보완적인 측면도 함께 지닌다. 전자는 CNN 내부의 특징 추출 과정에 색다른 경쟁 메커니즘을 도입함으로써 “채널 간 중복 최소화”와 “핵심 패턴 부각”을 달성하는 반면, 후자는 “주파수 성분의 정교한 반영”으로 ECG에서 흔히 발생하는 잡음 요소나 변동 폭에 대한 견고한 적응력을 제공한다. 결과적으로 본 연구에서 제안한 하이브리드 모델은, ECG 분석에 관한 기존 선행 연구들과 비교했을 때 정확도(Accuracy), 민감도(Sensitivity), 특이도(Specificity) 등의 주요 성능 지표에서 고른 양상을 보여주었다. 이는 단순한 구조 변경이나 하이퍼파라미터 조정으로 설명하기 어려운 의미 있는 개선이며, 본 연구가 제안한 모델의 독창성과 실효성에 대한 중요한 근거가 된다.

본 연구는 여기에 더해, 제안한 모델의 범용성과 실질적 확장 가능성을 검토하고자 두 가지 후속 실험을 병렬적으로 수행하였다. 첫 번째는 다중 클래스 분류 환경에서 MFCC 기반 모델의 적용 가능성을 평가한 것으로, 실제 임상에서는 단순 정상-비정상 이분법보다 다양한 부정맥 패턴의 판별이 요구된다는 점에서 출발하였다. 실험 결과, CNN+BiLSTM 구조는 다중 클래스 분류에서도 안정적인 성능을 유지하였고, 특히 MFCC가 적용된 경우 AAMI 기준 Q 클래스에서 더욱 두드러진 성능 향상이 나타났다. 이는 MFCC가 특정 클래스의 미세한 파형 차이를 더욱 뚜렷하게 부각시키는 역할을 할 수 있음을 의

미하며, 향후 클래스별 민감도 조정이나 주파수 기반 해석 연구로의 확장을 가능케 하는 실마리를 제공한다.

두 번째 실험은 MFCC의 핵심 구성 요소인 Mel filter bank 수(n_{mels})를 조정하는 방식으로 진행되었으며, 이는 입력 스펙트로그램의 정보량을 조절하여 모델의 과적합 여부 및 연산 효율성과의 균형을 평가하는 데 초점이 맞추어졌다. 실험은 n_{mels} 를 40, 64, 128, 256으로 다양화하여 진행되었고, AUC 및 F1-score를 중심으로 비교한 결과, n_{mels} 가 40일 때 가장 우수한 성능(0.982 / 0.937)을 보였으며, 이후 값이 증가할수록 성능은 소폭 하락하는 양상을 나타냈다. 이는 MFCC의 벡터 수(n_{mfcc})를 20으로 고정한 상황에서 필터뱅크 수가 과도하게 커지면 정보의 밀도가 낮아지며, 오히려 분류 성능에 불리하게 작용할 수 있다는 가능성을 시사한다. 또한 연산 효율성 측면에서도 n_{mels} 를 40으로 설정하는 것이 경량화와 실시간 적용 가능성 면에서 유리하다는 결론을 함께 도출할 수 있었다.

이러한 일련의 실험은 MFCC 기반 모델이 ECG 신호 분석에 있어서 일관되게 우수한 성능을 보이는 동시에, 설계 구성에 따른 민감도와 실용성을 함께 고려해야 함을 보여준다. 단순히 구조를 복잡하게 확장하는 것이 아닌, 정보의 핵심을 간결하게 추출하고 해석 가능한 방식으로 구성하는 것이 향후 생체 신호 기반 딥러닝 연구에서 더욱 중요해질 것이다.

본 연구에서처럼 통상적으로는 회피되던 중간 합성곱 계층에서의 Softmax 적용이 ECG 신호 분석에는 유효하게 작동할 수 있음을 확인함으로써, 모델 설계 과정에서의 새로운 관점을 제공한다. 다만, Softmax를 중간 계층에 도입할 때 발생할 수 있는 잠재적 문제점을 완전히 배제하기 위해서는 추가적인 정규화 기법이나 적절한 초기화 전략이 필요할 수 있다. 본 논문에서 사용한 데이터셋에서는 긍정적인 결과가 도출되었지만, 이 방법이 더욱 복잡한 형태

의 ECG 신호나 임상 현장에서 수집된 이질적인 데이터에 대해서도 동일한 수준의 성능을 유지하는지에 대한 면밀한 검증이 뒤따라야 한다.

아울러 MFCC 활용 역시, 음성 분석에서 주로 연구되던 기법이 생체 신호 분야로 확장되는 과정에서 개별 파라미터 최적화나 창 크기(Window Size) 설정, 필터 뱅크(Filter Bank) 구간 조정 등 다양한 설계 요소를 고려해야 한다. 본 연구에서는 ECG 파형에 맞춰 MFCC의 파라미터를 맞추는 과정을 실험적으로 진행했다. 그러나 이는 다른 바이오 신호나 다른 형태의 ECG 파형에 적용할 때 특화된 튜닝 과정을 거쳐야 할 수 있음을 의미한다. 따라서 MFCC 기반 하이브리드 모델이 보편적으로 활용되기 위해서는, 임상 환경과 다양한 환자군을 대상으로 한 추가 실험이 뒤따라야 할 것으로 보인다.

이런 방식의 본 연구의 시도는 CNN 계열 모델을 활용한 ECG 신호 분석에서 나타나는 대표적인 제약을 극복하려는 노력의 일환으로, 두 가지 핵심 아이디어 모두 충분한 가치를 지닌다. 첫째, MFCC를 통한 스펙트럼 기반 하이브리드 접근은 ECG 신호에 내재한 주파수 영역 정보를 깊이있게 반영함으로써 모델의 총체적 완성도를 높이고 둘째, 중간 합성곱 계층에 Softmax를 적용하는 방식은 기존 프레임워크의 틈새를 매우며 새로운 관점을 제시한다. 나아가 이러한 두 요소는 필요에 따라 독립적으로 적용하거나, 혹은 다양한 데이터셋 특성에 맞춰 보완적으로 확장할 수 있다는 점에서도 활용 여지가 크다.

결론적으로, 본 연구에서 제안한 중간 합성곱 계층에서의 Softmax 도입과 MFCC 기반 하이브리드 모델은 ECG 신호 분석 분야에서 새로운 가능성을 열어주었다. 전통적인 CNN 설계 흐름과는 다른 관점에서 출발했음에도, 실제 실험 결과로서 분류 정확도 향상과 잡음 저항성을 증명해 보였다는 점이 본 논문의 독창성을 뒷받침한다. 특히 추가 실험으로 진행된 결과에서 특정 부정맥(Q)의 경우 높은 분류정도를 보였고, 이를 바탕으로 먼저 처리된 Q값을 배제한 나머지 클래스들의 분류성능을 높이는 방향으로 임상적 실험을 이어갈

수 있다.

이러한 접근이 향후 더 다양한 규모와 복잡성을 지닌 ECG 데이터에 적용되어, 임상적으로 의미 있는 진단 보조나 실시간 모니터링 시스템 등에 기여할 수 있으리라 기대한다. 본 연구가 개척한 방향성을 이어받아, 후속 연구에서는 데이터셋의 범위를 확장하고 모델 하이퍼파라미터를 체계적으로 최적화하여, 실제 의료 현장에서 활용 가능한 수준의 안정성과 정확도를 확보하기 위한 다각적인 노력이 계속되기를 바란다.

향후 연구는 다양한 방법으로의 확장이 가능할 것으로 기대한다.

첫 번째로, 왜 softmax가 필터 활성화를 극단적으로 압축함으로써 ECG 신호의 특징을 도드라지게 만들었는지를 보다 직관적인 시각화 기법으로 확인해 보는 방안을 들 수 있다. 예를 들어 Grad-CAM, feature map visualization 등을 통해 “시계열의 어느 구간에서 어떤 필터가 활성화되었는지”를 살펴보면, softmax 기반 합성곱이 핵심 패턴을 어떻게 선택하고 강화하는지 명확히 이해할 수 있다. 이는 다른 생체 신호(EEG, PPG 등)나 음성 신호에도 확장 적용할 수 있는 잠재력을 시사하며, 다양한 실험 환경에서 재현성을 검증해나간다면 “특정 조건에서 softmax가 합리적인 대안이 될 수 있다”는 새로운 관점을 제시할 수 있을 것이다.

두 번째로, 분류 임계값(threshold) 설정에 따른 성능 변화 분석을 들 수 있다. 일반적으로 딥러닝 기반 이진 분류 모델은 0.5를 기준으로 양성/음성을 판단하지만, 임상 분야에서는 이 기준이 반드시 최적이라고 보기 어렵다. 특히 심장 질환과 같은 분야에서는 민감도(Recall, TPR)가 가장 중요한 성능 지표가 되기도 하며, 잘못된 음성 예측(False Negative, FN)은 환자의 생명과 직결되는 중대한 오류로 이어질 수 있다. 따라서 다소 낮은 특이도(Specificity)

를 감수하더라도 민감도를 높이기 위한 전략이 필요하다. 이와 같은 의사결정을 뒷받침하기 위해, ROC 곡선 분석을 통한 임계값 조정은 매우 효과적인 도구가 될 수 있다. 본 연구에서는 AUC 값과 f1-score를 주요 성능 지표 중 하나로 활용하였지만, 후속 연구에서는 특정 TPR-FPR 균형점에서의 성능 변화를 정량적으로 비교하고, 임상 목표에 가장 부합하는 threshold 값을 실험적으로 산정해볼 수 있다. 나아가, 환자 위험군에 따라 유동적으로 임계값을 조정하는 적응형 threshold 설계로까지 발전시킨다면, 모델을 단순 예측 도구에서 실제 임상적 의사결정을 보조하는 지능형 시스템으로 확장할 수 있는 가능성이 열릴 것으로 기대한다.

참고 문헌

- [1] Couceiro, J. P., J. P. Teixeira, and A. Lourenço. 2019. Classifying Heart Sounds Using Images of Motifs, MFCC and Scattering Coefficients with Convolutional Neural Networks. Porto: University of Porto.
- [2] Saadatnejad, Saeed, Mohammadhosein Oveisi, and Matin Hashemi. 2019. LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices. Tehran: IEEE.
- [3] Warrick, P., and M. N. Homs. 2018. Cardiac Arrhythmia Detection from ECG Combining Convolutional and Long Short-Term Memory Networks. Stanford: Stanford ML Research Group.
- [4] Oh, S. L., E. Y. K. Ng, R. S. Tan, and U. R. Acharya. 2021. Arrhythmia Classification Using BiLSTM with DTCWT and MFCC Features. Singapore: Springer.
- [5] Kiranyaz, S., T. Ince, and M. Gabbouj. 2019. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. Tampere: IEEE.
- [6] Hasan, M. K., M. I. Ibrahimy, and M. A. Hussain. 2016. ECG Signals Classification Using MFCC Coefficients and ANN Classifier. Kuala Lumpur: Springer.
- [7] Acharya, U. Rajendra, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. 2017. Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals. Information Sciences
- [8] 이경환, 황선명, 주영택, 변우용. 1994. 소프트웨어공학. 서울: 청문각.
- [9] Shi, Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-Chun Woo. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. Montreal: Advances in Neural Information Processing Systems. NeurIPS
- [10] Schuster, Mike, and Kuldip K. Paliwal. 1997. Bidirectional Recurrent Neural Networks. Brisbane: IEEE Transactions on Signal Processing.
- [11] Sokolova, Marina, and Guy Lapalme. 2009. A Systematic Analysis of Performance Measures for Classification Tasks. Montreal: Springer.
- [12] Moody, George B., and Roger G. Mark. 2001. The Impact of the MIT-BIH Arrhythmia Database. Cambridge: IEEE.
- [13] Davis, Steven, and Paul Mermelstein. 1980. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Transactions

on Acoustics, Speech, and Signal Processing 28(4): 357 - 366.

[14] Bengio, Yoshua, and Yves Grandvalet. 2004. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. Montreal: Springer.

[15] Liu, Zhongwei, Xiaoqiang Li, and Lei Wang. 2020. Automatic Arrhythmia Classification Based on Deep Feature Fusion and Weight Balanced Method. ResearchGate.

[16] Ma, Chi, Qingqing Cai, and Yanchun Zhang. 2021. Deep Learning for ECG Arrhythmia Classification: A Review. HeartSearcher Project. NIH

[17] Touluni, Youssef, Benayad Nsiri, and Belhoussine Drissi Taoufiq. 2023. ECG Signal Classification Using DWT, MFCC and SVM Classifier. Rabat: IEEE.

ABSTRACT

ECG Arrhythmia Classification Using a Hybrid CNN-BiLSTM: Task-Dependent Evaluation of MFCC and Softmax Effectiveness

This study proposes the application of Mel-Frequency Cepstral Coefficients (MFCC), commonly employed in speech signal analysis, to the analysis of ECG (Electrocardiogram) signals in order to address limitations found in existing deep learning-based approaches. Traditional methods based on FFT or wavelet transforms often suffer from temporal information loss or struggle to capture diverse waveform patterns. In contrast, MFCC offers the advantage of simultaneously reflecting nonlinear frequency characteristics and temporal features.

Hybrid deep learning architectures, including ConvLSTM and CNN+BiLSTM, were designed using MFCC-based inputs. Model performance was evaluated under varying conditions, including MFCC application, segment length (2s/3s), and activation functions (ReLU/Softmax) in convolutional layers. Notably, the experimental application of the Softmax function—unconventionally applied to convolutional layers—facilitated selective emphasis on salient features and contributed to improved performance. The CNN+BiLSTM model combining MFCC and Softmax maintained a consistently higher performance than models without MFCC.

In addition, the model's response to different input configurations and classification settings was investigated by modifying the number of MFCC filter banks and applying a multi-class classification scheme. By excluding Q-waveforms and focusing on clinically relevant classes, this study also explored the practical applicability of the approach. These findings suggest that MFCC-based ECG classification extends beyond structural enhancements and may contribute to real-time clinical decision-making and the development of advanced monitoring systems.

Key words : ECG, MFCC, Multi-class classification, Filter bank, CNN, BiLSTM, Softmax